AD-A056 760 HUGHES AIRCRAFT CO CULVER CITY CALIF F/G 5/5

A BEHAVIORAL MODEL OF TARGET ACQUISITION IN REALISTIC TERRAIN.(U)

JUN 78 L A SCANLAN, A K AGIN DAAK70-77-C-0013

UNCLASSIFIED HAC-REF-D8983 NL

1 OF 2

AD A056760

AD A056760

LEVEL II

12

B.S.

Report NO. P78-70R
HAC REF. NO. D8983

Prepared for:
VISIONICS Technical Area
Night Vision & Electro-Optics Laboratories
United States Army Electronics Command

FINAL TECHNICAL REPORT

A BEHAVIORAL MODEL OF TARGET ACQUISITION IN REALISTIC TERRAIN

JUNE 1978

AEROSPACE GROUPS

HUGHES

HUGHES AIRCRAFT COMPANY
CULVER CITY, CALIFORNIA

Approved for Public Release – Distribution Unlimited

(12)

Final Technical Report
P78-70R

A BEHAVIORAL MODEL OF TARGET ACQUISITION IN REALISTIC TERRAIN

L. A. Scanlan
and
A. K. Agin

Man-Machine Systems Section
Display Systems Laboratory
Engineering Division

and

Advanced Programs Laboratory
Tactical Systems Division

B.
use →

HUGHES AIRCRAFT COMPANY
Culver City, California
90230

June 1978

Prepared for

VISIONICS TECHNICAL AREA
NIGHT VISION AND ELECTRO-OPTICS LABORATORIES
UNITED STATES ARMY ELECTRONICS COMMAND

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER

2. GOVT ACCESSION NO.

3. RECIPIENT'S CATALOG NUMBER

9

6

4. TITLE (and Subtitle)
A BEHAVIORAL MODEL OF TARGET ACQUISITION IN REALISTIC TERRAIN.

5. TYPE OF REPORT & PERIOD COVERED
Final Technical Rep[redacted]t, 15 Dec [redacted]76 ~~to~~ 15 Oct [redacted]77,

6. PERFORMING ORG. REPORT NUMBER
D8983, P78-70R

10

7. AUTHOR(s)
L. A./Scanlan [redacted] A.K./Agin

15

8. CONTRACT OR GRANT NUMBER(s)
DAAK70-77-C-0013

9. PERFORMING ORGANIZATION NAME AND ADDRESS
Man-Machine Systems Section, Display Systems Lab., Engineering Div., and Advanced Programs Lab., Tactical Systems Div., Hughes Aircraft Co., Centinela & Teale Sts,, Culver City, CA 90230

10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS

11

11. CONTROLLING OFFICE NAME AND ADDRESS
Night Vision and Electro-Optics Laboratories (DELNV-VISD), Ft. Belvoir, VA 22060

12. REPORT DATE
Jun[redacted] [redacted]978

13. NUMBER OF PAGES
91

14

14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)
HAC-REF-D8983,
HAC-P78-70R

15. SECURITY CLASS. (of this report)
Unclassified

15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

12 102p.

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Target detection	Target acquisition model
Target search	Markov process
Target acquisition	Feature analysis
Scene Complexity	Visual Perception
Human Performance	

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

The research obtained eye fixation data while searching for targets in both realistic and abstract scenes. A Markov model of target acquisition is proposed and preliminary tests of its adequacy are made using the eye fixation data. The model considers the influence of input data, expectation, perceptual processing, and perceived scene information on the target acquisition process and offers considerable promise as a modeling approach.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

172 300

SB

ACKNOWLEDGMENTS

Contractual support for the reported research was provided by the Visionics Technical Area of the Night Vision and Electro-Optics Laboratories, United States Army Electronics Command, Contract Number DAAK70-77-C-0013. Mr. Richard Flaherty monitored the contract. The technical effort was conducted by the Man-Machine Systems Section of the Display Systems Department and managed by the Advanced Programs Laboratory, Tactical Systems Division, Hughes Aircraft Company, Culver City, California.

The authors wish to express their appreciation to the following people who contributed to the program and without whom the research could not have been accomplished. Mr. William Stewart, Program Manager, provided excellent and consistent leadership and wrote the Concluding Remarks section. Dr. Ronald Huss and Mr. Edward Dragavon developed the mathematical formulation of the Markov model, Mr. Clarence Dickson aided in the collection of the data and Mr. Tony Luchich provided technical assistance in the development of the analysis software.

A special thanks is extended to Dr. Michael Clark of Stanford Research Institute, Palo Alto, California, for allowing us to use his Perkinji Image Eye-Tracker. Dr. Clark's unselfish contribution of his time, energy, and expertise was critical to the successful collection of the experimental data.

ABSTRACT

Mathematical models of target acquisition performance, as representations of real world environmental situations, can be important and powerful tools for use in the design of electro-optical imaging sensor systems. To be useful, however, a model must predict accurately. This means that a large number of potential parameters and their interactions must be considered for inclusion in a complete model. Categories of parameters include: the characteristics of the sensor, the display, the atmosphere, the observer, the target, and the background. An examination of previous research and an analysis of the target acquisition process suggests a simplified two-component model as a basis for the development of a model capable of accommodating these parameters. The development of the two-component model and the results of two eye-fixation experiments which examine critical aspects of the model are presented in this report.

CONTENTS

CONTENTS (Continued)

LIST OF ILLUSTRATIONS

LIST OF ILLUSTRATIONS (Continued)

LIST OF TABLES

INTRODUCTION

Mathematical models of target acquisition performance are potentially powerful tools for use in the design of electro-optical imaging systems. The extent to which such potential is realized, however, depends upon the accuracy and generalizability of the predictions made by the model. The conditions encountered by operational sensor-display systems are highly complex, and an adequate model must include the characteristics of the sensor, processor, display, atmosphere, target, background scene, and the observer. The unusually large number of parameters which the model must ultimately include requires a logical and systematic approach to its development.

The present effort develops the foundation for such a model by integrating a diversity of research literature into a mathematical framework which is consistent with an analysis of the operator's target search and detection behavior. The general form of the resulting multi-component Markov model is outlined initially to provide an organizational basis for a discussion of the steps which lead to its formulation. Following a description of the psychological and physiological evidence in support of the model and a review of the mathematical aspects of a Markov process, the model is simplified for initial verification. Finally, the validity of the model approach is examined using data from two experiments which recorded eye movements during target acquisition. Because of the limited scope of the present research effort, an in-depth analysis of the experimental results could not be accomplished. A preliminary consideration of

the results are presented here with a wealth of pertinent technical data deferred for future consideration. The data do, however, demonstrate the validity of the approach selected and when fully exploited can be expected to provide a basic model with good predictive ability.

Why Model?

An adequate model of the performance of a sensor-display system as a function of the appropriate system parameters can be a powerful and cost effective tool for the system designer and the military strategist. The design engineer continually faces complex decisions in the selection of the optimum technical approach to be followed in the design of a new system. To make such decisions it is necessary to consider the technical advantage, the cost, and the expected system performance. With existing tools, most of the required data can be readily obtained or approximated. The performance of the operator using the system is the exception.

At the present time, data regarding the operator's performance must either be guessed at or evaluated empirically using either an actual system or a simulation of the system. Because of the time and expense of the empirical evaluation, the designer must rely on guessing or severely limit the number of alternative systems to be considered. In either case, the probability of an optimum or even near optimum decision is low.

Using a model, a designer can determine the impact of a large number of contemplated designs without the time and cost of building or simulating the systems for test. With the model implemented on an interactive computer terminal, for example, the engineer could input data regarding the characteristics of a candidate system, the manner in which it was to be used, and the anticipated cost. The computer could use this data and the detection model to compute and output the performance increment per dollar cost. In a matter of hours a large number of alternatives could be explored and the optimum configuration retained for further study. Without the model, the examination of the same alternatives could require many years of effort.

If the model also reflects the characteristics of the underlying behavioral mechanisms, it is possible to identify those aspects of the system and task which are the most difficult for the observer. Such knowledge provides the necessary information to allow effort to be concentrated on those parameters or procedures which will result in the largest increase in system performance. As an example, knowledge of the influence of the displayed scene content on observer performance could provide invaluable direction to the development of highly effective and efficient image processing techniques. A strategist might also exercise the model with parameters representing an existing system under a variety of tactical situations to assess the best method of deploying the system.

Modeling Approaches

As with any complex problem, more than one approach can be taken to the development of a mathematical model of target search and detection. The optimality of an approach depends upon the ultimate goal of the model. If the goal is to obtain a short term model which predicts to a limited set of specific conditions then a data fitting procedure may be the best approach. On the other hand if the goal is to ultimately obtain a model capable of performing the functions outlined previously, then an alternative approach may be required. These alternative approaches are explored below.

Equation Fitting Approach. Most existing models of target acquisition have used data fitting approaches to obtain equations which predict the probability of detection as a function of time. These models are generally based on a Poisson process(24) of the form $P(t) = P_\infty(1 - e^{-t/\tau})$ because the shape of the resulting curve is similar in shape to the observed probability of detection as a function of time. The parameters P_∞ and τ depend upon certain system parameters and are empirically determined using curve fitting techniques. Models of this type have been successful in predicting performance for abstract targets(72) and for simple or uniform background

conditions.[34,65] However, these models do not generalize to the prediction of performance with realistic, complex backgrounds.[54]

Because the number of potential scene characteristics can be large and their effects varied, the two parameter model described above may not be capable of describing the observed behavior. For example, Figure 1 plots the cumulative probability of correct detection as a function of time for two of the conditions from Scanlan.[54] As can be readily observed, the shape of the cumulative probability curve for the more difficult high complexity scene, low target-to-background contrast condition is much different than the curve for the low complexity scene, high contrast condition. Three parameter models such as $P_\infty (1 - e^{-(t/\tau)^k})$ or $P_\infty (1 - e^{-t/\tau})(1 - e^{-t/\tau_2})$ may be necessary. This latter form will result in a cumulative probability function similar to the top curve of Figure 1 as τ_2 approaches 0 but will resemble the lower curves as τ_2 departs from 0.

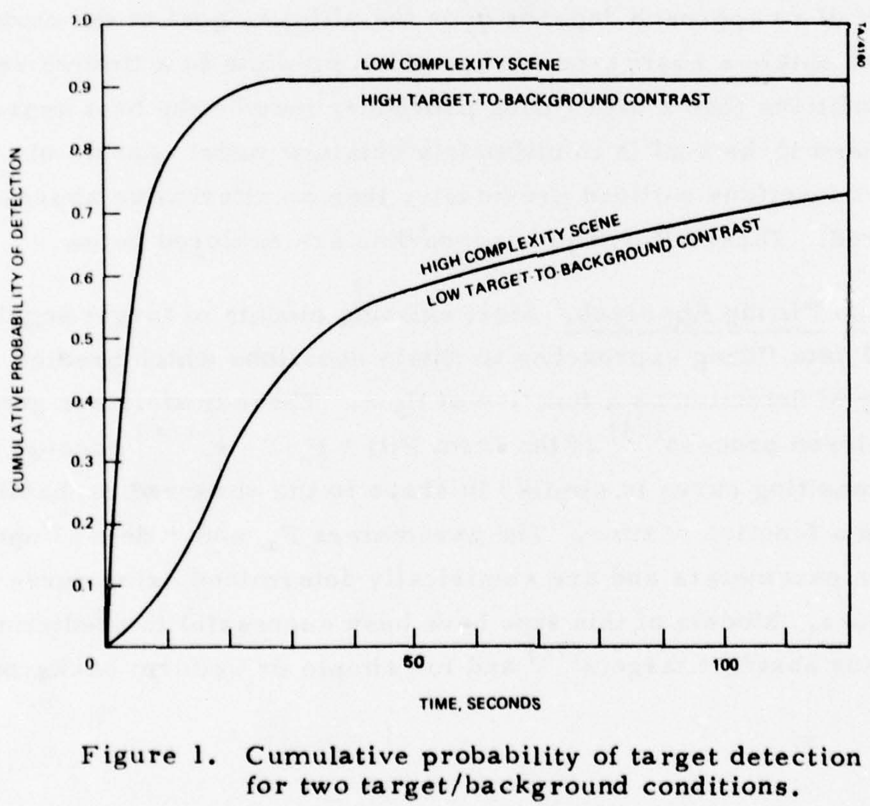


Figure 1. Cumulative probability of target detection for two target/background conditions.

In addition to a change in the equation form, a tremendous amount of new data would be required if the final equation is to predict to a large number of conditions. A list of the variables and parameters which conceivably might influence performance, and hence, should be included, rapidly becomes very long. Easily 50 to 100 variables might be considered. If each were examined at only five levels, an incomprehensibly large quantity of data would be required to examine all possible combinations of each variable at each level.

Extensive use of screening studies[61] and considerable expert judgment might reduce the number of variables to 25. Assuming the judgments were all correct, data could be collected on the reduced set of variables and used to fit the model. Further, if it can be assumed that most of the variance of interest is in the main effects and second order interactions, Response Surface Methodology[11, 12, 16, 60, 74] sampling procedures could again reduce the data requirements by several orders of magnitude. Even with these reductions, the total number of observations, assuming five levels of each variable, would still be approximately 3 quadrillion or 3 million million.

Imagining for the moment that such data, or a subset, could be obtained, the resulting equation would only be an approximation and would be limited to the specific conditions examined with a very low probability that the relationships obtained would hold for any case not specifically examined. Thus, each new technological development could require a new and massive data collection effort, because it is not possible to know or even guess with any certainty the performance to be obtained from a system not examined in the derivation of the original model. As a result of the gross inefficiency of the direct observation and curve fitting methods and the inability of the resulting model to accommodate changing technology, missions, and environments, those methods are woefully inadequate.

An Alternative Approach. The equation fitting approach assumes that little or nothing can be known about the human operator and the underlying causes which lead to the observed behavior. This assumption is false. The human is highly complex and is capable of a wide diversity of response to what often appears to be identical situations. This, however, does not mean that the human must always be considered as a black box whose inner workings are never to be fathomed. On the contrary a great deal of psychological and physiological evidence has been obtained which provides an insight into the manner in which information is processed in the human perceptual system.

An application of the knowledge concerning human perception can be a powerful method of reducing the large number of potential variables to a few critical ones which reflect the information used by the observer when engaged in a search for a target. If variables such as target-to-background contrast, sensor and display resolution, field-of-view, target type, and scene characteristics, to mention only a few, could be integrated into a few underlying informational content variables, the problem of developing a model could be greatly simplified. These system parameters affect the amount and type of data from which the operator may extract relevant information. The operator does not particularly care how the data was obtained but only that it contain the required information. Because the operator is concerned about information, variables which describe that information should be the appropriate ones for a model of operator performance. The many system variables would then be transformed in terms of their affect on the information available to the operator.

If these underlying variables also reflected the characteristics of the observer, then it would no longer be necessary to obtain new data each time technology produced a new system capability. The characteristics of the new system would merely be cast in terms of the underlying variables and the model exercised with these inputs. It may still be desirable to confirm the predictions of the model; however, only a few specific

conditions need to be examined to insure that the model continues to predict to the new combinations of conditions.

A modeling approach which considers the information processing characteristics of the human perceptual system is pursued in the present effort. This more general and more powerful approach is based on an analysis of the operator's task and integrated into a mathematical framework with clearly defined and *testable components*. The model provides a means for appropriately representing and incorporating both the processing states and the sources of information characteristic of the target search and detection task. The multi-component model and the rationale behind it are briefly outlined in the next section. This brief description is followed by a detailed examination and integration of the available literature which lead to the model.

Multi-Component Model

Initial evidence for a multi-component model was based on a comparison of the operator's performance under uniform background conditions and under complex realistic conditions.[54] On a static monochrome display with a stationary target, detection and recognition can be accomplished along only two dimensions: luminance and spatial. In the case of a target located in a uniform background, the luminance factors will predominate, because there is no need to discriminate shape characteristics of the target. However, a target located in a real-world background must be detected on the basis of both luminance and spatial characteristics.

The shift in relative importance of spatial cues can be seen in Figure 2 which presents the interaction of displayed resolution and background type on the time required to detect a vehicle target. With a uniform background, display resolution had no effect, while with realistic backgrounds, a variation in display resolution resulted in a better than two-to-one change in time to detect. Clearly, the importance of spatial detail was minimal when the target was located in a uniform background but was of considerable importance when it was necessary for the target to be discriminated from conflicting objects with similar characteristics.

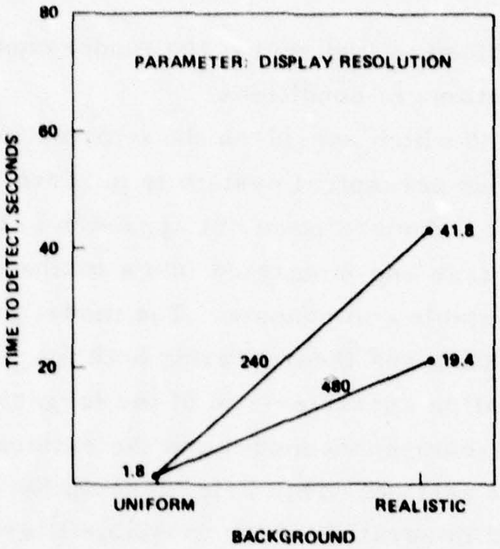


Figure 2. The effects of background and display resolution on the time required to detect a target (from 54)

This interaction of background type and display resolution can be interpreted as evidence for a multi-component model of target search and detection.[54] Data from other studies also lend support to this interpretation.[72, 40] Although, the hypothesis requires further examination, such a multi-component conceptualization does reconcile much of the observed data. Further, it separates a complex task into behaviorally meaningful parts which can be investigated individually to determine the effect of the input data and the task to be performed. The potential benefits of the approach are sufficient to justify further elaboration of the model and an investigation of its accuracy.

A multi-component model of target search and detection, based on an analysis of the task and existing literature, is graphically presented in Figure 3 as a Markov Process. The model includes four processing components or states which have a number of transitional probabilities (q_{ij}) associated with them. Two terminal absorbing states represent the

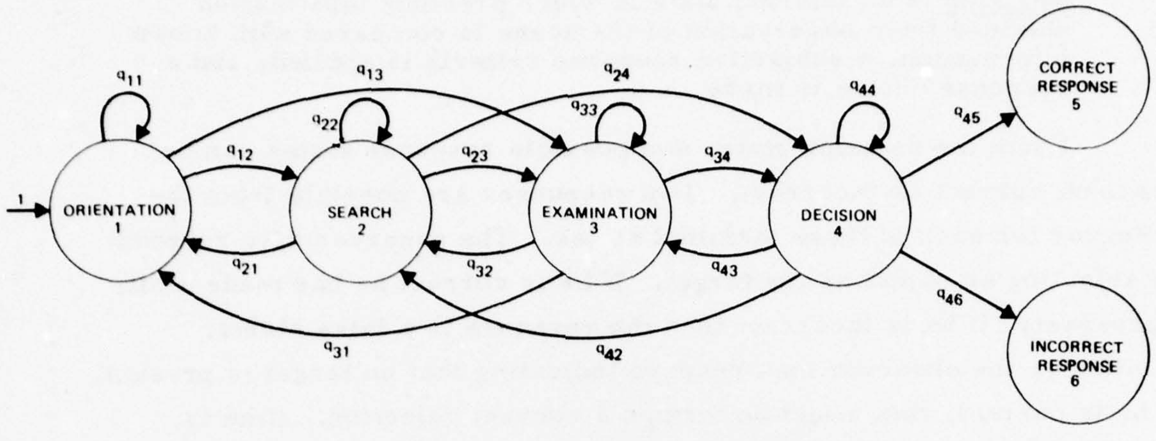


Figure 3. Multi-component Markov process model of target acquisition.

final response outcomes. As shown, the process always begins in the orientation state, and re-enters the same state or transition to another state at fixed time increments. Transition probabilities are a function of the large number of variables that can affect detection performance and are assumed to remain constant over time. The cumulative probability of target detection as a function of time is obtained by considering all the possible paths to the correct response state and the time required to traverse those paths.

The four processing states are called orientation, search, examination and decision, and are defined as follows:

> Orientation is an initial processing of large and salient characteristics of the scene. Under realistic conditions, orientation would consist of a brief, wide-angle look at dominant terrain features such as roads, trees masses, lakes, and fields which form meaningful patterns and gross relationships. These result in a global search strategy.
>
> Search is the processing state in which sub-areas of the scene are examined by short eye fixations on objects likely to be targets.
>
> Examination is the processing state in which an object selected as a potential target is scrutinized in greater detail to determine if it is a target.

Decision is an internal state in which previous information obtained from observation of the scene is compared with known information, a subjective response criteria is applied, and a response choice is made.

From the decision state, two possible response states can be reached: correct or incorrect. Two responses are possible from the observer for each of these terminal states. The observer may respond by selecting an object as the target. If he is correct he has made a hit. Conversely, if he is incorrect then the response is a false alarm. Similarly, the observer may respond indicating that no target is present. If he is correct, this would be termed a correct rejection. If he is incorrect, the outcome would be a miss.

As illustrated in Figure 3, the multi-state Markov model is highly adaptable to the complex target acquisition situation. Given that states are defined with valid and invariant characteristics, the model provides for alternative paths and sequences such that states may be repeated, skipped over, or entered in varying temporal order. The model also allows for expansion by systematically expanding individual states into sets of sub-states. For example, search is a likely candidate for expansion into a set of sub-states representing search within specific types of scene areas. In an expanded model, the set of transition probabilities into and out of the sub-states would replace the overall transition probabilities into and out of search as a whole. The validation of the basic component states, followed by an expansion of each state systematically provides a logical approach necessary for the development of a model of target search and detection.

General Approach

In the following sections, the several steps in the development of a preliminary multi-component Markov process model are presented. First, a review of the available literature on perceptual processing is combined with an analysis of the operator's task to provide a framework

which relates the search and detection task to measurable and behaviorally meaningful variables. Second, a simplified two-component model is developed for initial experimental evaluation. Third, two experiments which used operator eye fixations on the image as a method of measuring the probability, sequence, and duration of the processing components are detailed. Fourth, a preliminary evaluation of the model and approach is presented which confirms the validity and strength of the approach. Finally, directions are suggested for future research and model development.

HUMAN INFORMATION PROCESSING AND TARGET ACQUISITION

The perceptual information on which operator performance ultimately depends is not a direct function of the data input through the sensory mechanisms of vision. Rather, it is a complex function of the input data, processing mechanisms, and operator expectation as indicated schematically in Figure 4. Because the perceptual information used by the operator represents a highly processed and transformed subset of the total input data, the development of a model capable of adequately predicting performance in realistic situations must begin with a consideration of the perceptual information relevant to target search and detection.

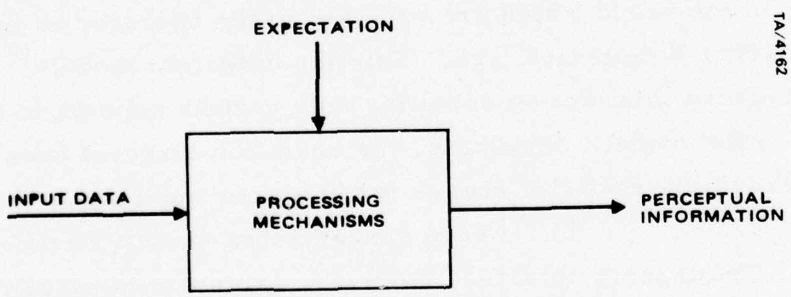


Figure 4. Aspects of perceptual data processing.

An identification of relevant information requires a consideration of both the processing capabilities of the operator, as they are currently understood, and the task. Further, because the processing required to obtain the necessary perceptual information changes with the expectation of the operator, this internal source of information must also be considered. In the paragraphs which follow, input, expectation, and processing characteristics will be examined in the context of the target search and detection task to yield hypotheses concerning the relevant perceptual

information. These hypotheses will then be integrated into the multi-component Markov model.

Input Data

A portion of the visual world as it impinges upon the eye is the input data. More precisely it may be defined as the objective physical temporal and spatial variation in luminance intensity. In the most general sense the input data is the total data required to reconstruct a physical real world situation and will be a function of a large number of factors. If a sensor-display system intervenes between the world and the observer, the input data becomes a function of the sensor, display, and sensor-to-display transformations as well as the characteristics of the objective world. A complete description of the world, obviously, requires an incredibly large quantity of data and only recently have serious attempts at a description been considered.[22]

Although a complete description of the world might be possible, it may not be necessary. Because the operator processes the input data to extract a sub-set of the original data, it may be adequate to describe only those aspects of the world which are relevant to the operator as he performs the target search and detection task. Existing detection models[9,23,34,65] implicitly recognize this fact by selecting only certain aspects of the world for inclusion in the model. Similarly, the research directed toward the problem of describing realistic scenes for operator modeling and computer recognition[3,4,22,37,41,46,69] have concentrated on only certain aspects of the scene. The aspects selected, however, are not necessarily those of importance to the observer. Clearly the requirement is for an understanding of the perceptual information extracted from the input data. This requires an examination of the processing capability of the operator.

Perceptual Processing

Evidence from both the neurophysiological and psychological literature make it clear that the perceptual system extracts features from the input data and that these features are the building blocks from which perceptual information is constructed.[7,18,19,31,32,33,42,47,51,59]

A review of these studies is beyond the scope of the present effort. However, it is generally agreed that features such as intensity, size, color, retinal location, orientation of long axis, spatial frequency, acute angle, line length, and motion are extracted from the incoming visual stimulus.

Physiological evidence also exists which suggests a hierarchy of feature processing.[30, 35, 52] Low-level features such as size, contrast, color, and retinal position are relatively automatic in their coding, require little processing time, and are extracted at the more peripheral areas of the visual systems. High-level features which may be composed of several components, such as edges and angles, usually have high spatial frequency components,[44, 56, 57] require foveal processing, and increased time for their extraction.[36] These high-level features are most likely extracted at the visual cortex.

The extraction of features from the input reduces the amount of data dramatically and codes it in a more useful form. However, the feature set requires additional processing to become perceptual information. The features must be selected, ordered, weighted, and combined in a meaningful way. The logical basis for constructing perceptual information from features is a function of expectation.

Expectation

The information available to the operator through memory is called expectation. Expectation is a function of the experience and normal perceptual development of the operator, as well as specific briefing before the task. It can be reasonably assumed, on the basis of past research, that the population of normal adult operators shares common rules for decoding realistic scenes. These include: perspective rules that map three-dimensional objects onto two-dimensional displays; segmentation rules which separate discrete objects from the background; and relational rules for arranging objects within realistic scenes.[13, 22, 37, 43, 53] Briefing adds specific information about the target, terrain, mission, sensor and display, and response criteria within the specific target acquisition task.[34, 55, 58]

The perceptual system uses expectation and the feature description of the input data to construct perceptual information. The stimulus data will be used to the greatest extent possible consistent with expectation which governs the construction process. If stimulus information is inadequate, the construction process will supplement the stimulus material to provide a reasonable organization. If the input data is conflicting when considered in the context of the subject's expectation, then some data will be rejected or distorted.

Perceptual Information in the Scene

An identification of categories of information within a complex, realistic scene can be obtained by asking observers to describe those characteristics of the scene which might make target detection easy or difficult. On the assumption that subjects share common expectations in this situation, the objects identified can be expected to represent the categories of potentially relevant perceptual information.

In a previous research program,[54] 12 subjects provided opinions as to which scene or target characteristics would aid or hinder detection of the target. A distillation of those responses yields four categories of scene information: target, clutter, context, and texture. Each of these can be described in terms of feature characteristics and expectations.

Target. The target will have a set of perceptual features similar to any complex visual pattern.[10, 36, 43, 44, 68, 78] Considerable target acquisition research has been directed toward identifying relevant target features with emphasis on the low-level features of size and contrast.[9, 34] Under realistic background conditions, these features may not be sufficient. Detection is considered a correct response indicating that an object is located in the scene; however, realistic backgrounds introduce many non-target objects, requiring additional discrimination between target and non-targets. Therefore, even in detection tasks, the target must have some high-level features which distinguish it from non-targets. The number and type

of features processed under these conditions requires further study. It would be expected, however, that characteristics such as specific shape or outline, internal detail, and internal modulations are likely features.[54]

Clutter. Clutter is, collectively, those objects which are detectable and share some features with the target. In general, clutter and a target will have similar low-level features but will differ on some high-level feature characteristics. The number of common features between target and clutter, the proximity of clutter to the target, and the number of clutter objects are factors which can influence detection performance.

The number of clutter objects in controlled abstract studies can account for as much as 97% of the variance in the data.[45] Significant effects of number of clutter objects on time to detect[2] and probability of detection[73] have been found in target acquisition studies. The number of clutter objects is used as a background parameter in several target acquisition models including GRC and MARSAM models;[65] however, the density, or number of objects per unit area may be a better predictor of performance.[54] Clutter objects close to the target have the largest effect on performance.[9, 34, 73]

Similarity is a function of the number of features shared by clutter and target[20] and can have a major effect on performance. Low-level features such as size and contrast are the most commonly measured features.[8, 14, 29, 34, 62] Objects having a size range up to four times the actual target length are likely to affect performance.[34] Shape, or high-level features have not received as much systematic investigation, particularly for tactical vehicle images. In general, the shape features which have been studied include a geometric category (circle, square, rectangle)[45, 62, 72, 73], and the number and relation of linear and angular components.[35, 44]

Context. Context relates to those terrain objects or areas which have a systematic and meaningful relationship with target location. For example, roads are context objects with a very strong functional relation to tactical vehicles, and detection may be a function of the proximity of the target to roads. Roads and most other terrain objects are usually of low spatial frequency and are not likely to be confused with targets. Normally, context objects seem to facilitate the search process, because they offer the operator information about areas where targets are likely to occur, and also areas where targets are not likely to occur, thus, indicating areas the operator should reasonably search or ignore. Context objects may have widely varying photometric characteristics but their effect on target acquisition is a function of the logical rules related to expectation.

The concept of context has been included in some analyses of target acquisition.[9,39,40,73] Location constraints or areas likely to contain targets affected probability of detection[73] and areas likely to contain targets interacted with amount of clutter in predicting total time to detect.[54] *Other studies have reported* very little difference in performance as a function of context. Whittenburg, Schriber, Robinson and Nordlic[71] did not find differences in performance when targets were placed in open areas as opposed to targets placed near terrain features. Krebs and Graf[40] found only a small trend in performance with respect to percent usable area within the scene.

These conflicting results argue for further clarification of the concept "areas likely to contain targets," and the type of context features which have the greatest predictive effect on performance. As an example, a tactical vehicle such as a tank has very clear performance limitations with respect to terrain.[67] If the tank's grade ascending limit is 60 percent, areas with steeper grade may be eliminated from search. If the tank has limited fording ability, it is not likely to be found near water. The performance characteristics of the target are well-known to experienced personnel and contribute to the observer's expectations.

The effects of pattern and masking can be included in context. It has been found with eye fixation prediction models that certain geometric patterns are more likely to be fixated than others. For example, ends of straight lines, vertices of acute angles, and intersections of straight lines are very likely to be fixated.[49, 50, 64, 77] Targets placed with respect to various geometric patterns created by roads, treelines, rivers, etc., may be detected partially as a function of such pattern effects. While not true logical context effects, geometrical patterns have been included, because they relate to terrain features. It would be predicted that acute angles, right angles, obtuse angles and straight lines will elicit fixations and will, therefore, have significant effects on detection if targets are located near these terrain features. Open areas, irregular surrounding areas, and locations contrary to geometric pattern effects should not elicit fixations and may result in a performance decrement. The masking of targets by terrain features which occurs when the terrain is viewed from low altitude can also significantly affect performance.[21, 24, 34]

The target acquisition literature suggests several candidate type of context objects. Roads, open fields, bridges, lakes, forest edges, and man-made objects have been considered.[15,34,54,55,58] The research on the automatic recognition of terrain features such as fields[3] and overall terrain type[69] may ultimately provide additional insight into the characteristics of context.

Texture. Texture refers to areas with relatively uniform or recurrent elements of high spatial frequency and low modulation.[48, 54] The effect of texture on performance is expected to be relatively small, and will not be considered further at the present time.

Scene Information and Target Acquisition

The previous discussion considered the influence of object features and expectation on the perceptual information in the scene. These aspects of perception can be related to the four states of target search and detection

given in Figure 3. The four stages - orientation, search, examination, and decision - are states in a Markov process which may be entered sequentially or iteratively.

Orientation. The initial input of low-level, global scene features with rapid eye fixations over a large area of the scene characterizes orientation. The average eye fixation for an operator scanning a display is approximately 300 msec. The rapid scanning over a wide area which is typical of picture processing at the beginning of a task[1] is an indicant of orientation. The outcome of orientation is a simple global scene description and a general strategy for searching selected areas. Expectation and context are predicted to be the major influences on orientation. Although orientation determines much of the sequence and probability of fixations during search, it is expected to be relatively brief and constant across conditions.

Search. The rapid fixation of areas and objects in the scene according to the search strategy developed during orientation is called search. Objects are briefly fixated during search and low-level features are extracted. Objects having a number of low-level features in common with the target will elicit a transition to examination; objects not having the relevant low-level features result in a continuation of search. The number and type of clutter objects can be expected to have a major influence on search. Because the duration of each fixation is only 200 to 400 msec, feature extraction is limited to low-level features.

Examination. The high-level features of candidate target objects found during search are extracted in the examination state. If the candidate target object has the relevant high-level features, a transition to the decision state occurs. If the high-level features are not found, search is resumed. Fixations in this processing state are confined to a small area around the candidate target, and the total duration of these fixations will be longer than the nominal 200-400 msec characteristic of search.

Decision. The high-level features extracted during examination are interpreted according to expectation and a decision to make a response or continue search or examination is made. If the decision is to respond, then the selection of the appropriate response is also made in this state.

Mathematical Model

One obvious and well-developed mathematical representation of a process involving probablistic transitions between states is a Markov model. This model characterizes systems where the conditional distribution of the random variables is independent of the past history of the system. That is, the system does not "learn." This characteristic manifests itself in this model in the assumption that the transition probabilities from state to state do not change as a result of the past behavior of the system. Although this assumption may not be strictly true, it provides a reasonable starting point, and any inadequacies will become evident as the model is exercised in an experimental setting.

The Markov model is also attractive in that it can be treated at varying levels of complexity. The simplest approach, a first-order Markov process, is to assume that transitions from one state to another occur randomly and follow a Poisson distribution. Such an approach is shown in a subsequent section to result in a linear differential equation.

The Markov model is, however, capable of handling more complex processes. Varying levels of "memory" can be introduced, such that the transitions can be functions of the state of the system for the previous n state occupancies, where n can be any chosen value.

The end product of this model is a cumulative probability of a correct detection as a function of time. This distribution can be obtained from the model in various ways. The first-order model, possessing an analytic solution, can be solved directly as a continuous function of time. Alternatively, the solution can be discretized in time, with the state of the system at time $t + \Delta t$ being determined by multiplying the state vector at time t by the transition probability matrix. These evaluations will necessitate the use of a digital computer.

VERIFYING THE MODEL

To test the model, it is necessary to estimate the Markov model parameters on the basis of operator behavior as a function of task variables. The monitoring of eye fixations on the scene during target search and detection is an appropriate method. The use of eye fixation measures is supported by results reported in the research literature, and such measures provide the necessary duration, sequence and probability estimates required to compute the Markov model parameters. In this section, the properties of eye fixations will be reviewed, the modifications of the multi-component Markov model into a two-component testable format will be presented, and the plan for two experiments will be summarized.

Eye Fixations

It has long been established that individual eye fixations with respect to specific areas of the scene are good indicators of the information which is sampled by the operator during target search and detection.[1,20,27,41,49,50,72] Eye fixations are a function of the size and contrast objects,[20,73] and their relative importance with respect to the task.[1,26,75] Targets elicit more frequent fixations than non-targets, and complex targets more than simple ones. Often in target acquisition tasks, the target is re-fixated several times before a response is made.[40] The number, type, and variety of clutter objects in the background also affect fixations. For example, if clutter objects are all different, more fixations are required to detect a target than when clutter is essentially homogeneous.[45]

Overall distributions of eye fixations over a variety of scenes demonstrate some consistent trends that are independent of scene content. For example, eye fixations are more widely-spaced the larger the display size and the closer the viewing distance.[40,70] There is also a trend toward fixating the center of a display rather than the edges;[9] and the lower third of the display which represents closer objects when real terrain is displayed.[27]

Fixations are not uniformly distributed over a realistic scene during target acquisition. Even when the task is only to look at the scene,[49] certain areas with high information content are selected for fixation and other areas are ignored. The nature of the task and the immediately observed characteristics of the target strongly influence the fixation pattern.[75] For air-to-ground target search and detection, Snyder[63] reported that from 80 to 90 percent of all fixations were restricted to 5 percent of the available scene area. The selectivity of the eye fixation patterns supports the view that eye fixation measurement may serve as a method of determining the type of information that is processed and may be applied in a model of processing components.

Eye fixations have several characteristics which may be applied to advantage in a Markov process model. They vary spatially as a function of information. Fixations also vary in their temporal characteristics such that the sequence, duration, and frequency of occurrence may also be measured. Although the instrumentation required to measure eye fixation variables has been, until recent years, extremely difficult to obtain, accurate point of regard systems are now available.

Another characteristic of eye fixations which is potentially important in modeling target search and detection is the size of the information input area that is active with each fixation. It is generally understood that the foveal lobe area subtending about 1 to 2 degrees diameter about the visual axis is the primary source of information input from an individual fixation. However, through selective attention and varying information demands, the actual Data Input Area (DIA) at any point of time, may change significantly.[4,24,49,76] An average DIA, or lobe area is employed in this initial modeling effort, and it is the lack of DIA assessment methods

which requires an initial simplification in the multi-component Markov model.

Simplification of the Model

The multi-component Markov model includes four processing states: orientation, search, examination and decision. Using eye fixation measurements, it is possible to operationally define each processing state. These operational definitions then direct the modifications that should be made in simplifying the model.

The two states which are difficult to assess experimentally are orientation and decision. The operational definition of orientation would include a relatively short average duration and a very large lobe or DIA. Orientation is by definition the state associated with the first fixation, but subsequent fixations may also be in that state. Since there is no easily implemented method of measuring DIA, orientation at present should be combined with search. Orientation and search share a short average duration characteristic; and orientation is relatively short and constant in its overall contribution to performance.

Decision is also difficult to measure, since it includes by definition, the internal process of response selection. When the decision to be made is relatively simple and the number of alternative responses are small, the decision process will be short and its contribution to performance relatively constant thereby allowing it to be combined with examination.

The simplified model is illustrated in Figure 5 as a two-component Markov process model with two terminal absorbing states representing correct and incorrect responses. The model includes a search state and examination state, defined by their different temporal and informational characteristics. Search is characterized by short duration fixations while examination fixations are longer. Examination fixations include repeated fixations to an area, as well as clusters of sequential fixations in a common area. Overall statistics of eye fixation duration do not necessarily demonstrate the two component distributions. The model must be tested by determining where the fixations occur, and whether individual fixations are short or long.

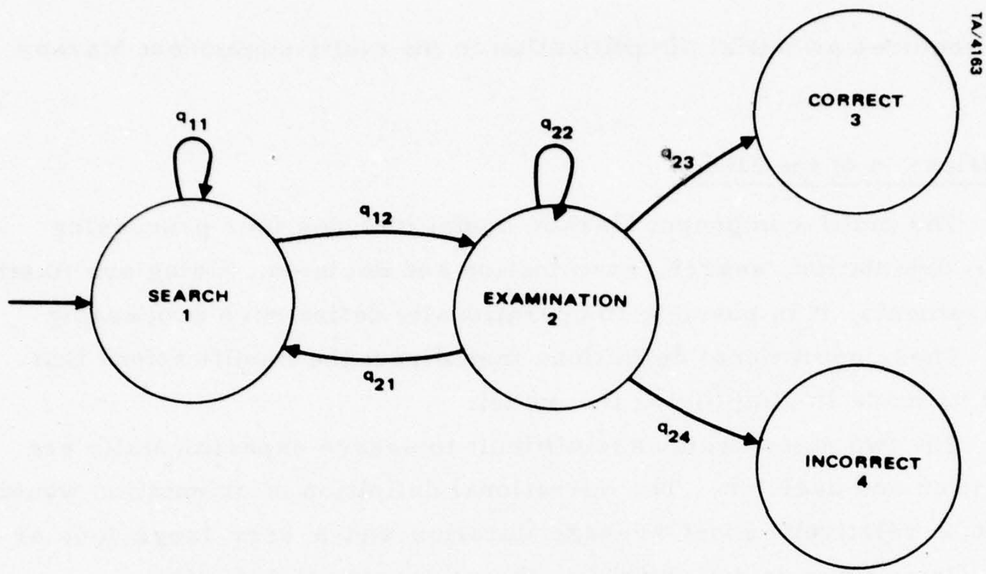


Figure 5. Simplified Markov Model of target search and detection.

Markov Equations

The derivation of the mathematical model of the Markov process is straight-forward. Figure 5 represents a four-state, first order process in which the probability of a transition from state i to state j is denoted by q_{ij}. It is assumed, at this stage of the modeling effort, that the probability of a transition from one state to another is a Poisson process. In a Poisson process, where δt is small, the probability of a transition in a time interval δt is proportional to δt. Thus, if the system is in the ith state at time t, the probability of jumping to another state at time $t + \delta t$ is $\delta t/\mu_i$, where μ_i is the mean duration spent in the ith state.

If a transition occurs, it will take one of several paths, represented by q_{ij}. The probability that the system will be in state j at time $t + \delta t$ is thus $1/\mu_i \ \delta t \ q_{ij}$. Now let $x_i(t)$ be the probability that the system is in state i at time t. It will still be in state i at time $t + \delta t$ if it does not undergo a transition (with probability $1 - 1/\mu_i \delta t$) or if it was in another state j and has a transition into state i. The latter is expressed by

$$\delta t \sum_{\substack{j=1 \\ j=i}}^{N_1} \frac{1}{\mu_j} q_{ji} \, .$$

Writing these expressions for all systems states leads to the expression for the state probability function at time $t + \delta t$, namely,

$$\begin{bmatrix} x_1 (t + \delta t) \\ x_2 (t + \delta t) \\ \vdots \\ x_n (t + \delta t) \end{bmatrix} = \begin{bmatrix} 1 - \frac{1}{\mu_1} \delta t & \frac{1}{\mu_1} \delta t q_{21} & \cdots \\ \frac{1}{\mu_2} \delta t q_{12} & 1 - \frac{1}{\mu_2} \delta t & \mu_2 \delta t q_{32} \\ \vdots & \vdots & \ddots \end{bmatrix} \begin{bmatrix} x_1(t) \\ x_2(t) \\ \vdots \\ x_n(t) \end{bmatrix}$$

In the limit, as δt goes to zero, this becomes the differential equation

$$\frac{dx}{dt} = Ax(t), \quad t_o \le t,$$

where the coefficient matrix A is given by

$$A = \begin{bmatrix} -\frac{1}{\mu_1} & \frac{1}{\mu_1} q_{21} & \frac{1}{\mu_1} q_{31} & \cdots \\ \frac{1}{\mu_2} q_{12} & -\frac{1}{\mu_2} & \frac{1}{\mu_2} q_{32} & \cdots \\ \vdots & \vdots & \ddots & \end{bmatrix}$$

In this derivation no mention was made of the characteristics of the parameters μ_l or q_{ij}. Although these may be (and probably are) functions of time, considerable experimental data would be required to determine the form of such a variation. A first approach toward fitting the experimental data would be to assume that the μ_i and q_{ij} remain constant over all t; under this assumption the differential equation has the solution

$$X(t) = X_o e^{A(t - t_o)}, \quad t_o \leq t,$$

with $X(t_o) = X_o$. This is probably the simplest mathematical representation of the model that retains the integrity of the parameters.

This differential equation can be approximated numerically with the aid of a digital computer. An initial state vector (X(o)) is entered and multiplied by the transition matrix containing the μ_i and q_{ij}. The vector resulting from this is again multiplied by the transition matrix to determine the state probability vector after two time steps. This process is continued until some correct decision probability criterion is reached. The solution generated in this way contains a record of the probability that the observer will be in each state at any time t.

Specifically, the continuous time Markov process may be approximated by an equivalent discrete time process as follows: Let |S| be a row vector of probabilities of state occupancy.

$$|S| = |S_1, S_2, S_3, S_4|$$

and let |T| be the matrix of transition probabilities where row indices represent the current state and column indices represent the state at time k+1.

$$|T| = \begin{vmatrix} q_{11} & q_{12} & q_{13} & q_{14} \\ q_{21} & q_{22} & q_{23} & q_{24} \\ q_{31} & q_{32} & q_{33} & q_{34} \\ q_{41} & q_{42} & q_{43} & q_{44} \end{vmatrix}$$

Then |S| at k+1 will be equal to |S| at time k time |T|.

$$|S|_{k+1} = |S|_k \, |T| \; .$$

For the Markov process of Figure 4, |T| simplifies considerably because a number of the transition probabilities are either 1 or 0.

$$|T| = \begin{vmatrix} q_{11} & q_{12} & 0 & 0 \\ q_{21} & q_{22} & q_{23} & q_{24} \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{vmatrix}$$

The remaining transistion probabilities can be determined from the eye position data

1. Divide the time base into arbitrary intervals.
2. For each interval determine the state.
3. Count the number of transitions from state i to state j, for all i and j assuming transitions only occur at end of each time interval.
4. Compute q_{ij} by dividing the number of transitions from state i to j by the total number of transitions from state i. For example

$$q_{12} = \frac{N_{se}}{N_{ss} + N_{se}}$$

where

N_{se} = number of search to examination transitions

N_{ss} = number of search to search transitions

The cumulative probability of a correct response as a function of time can be obtained by iteratively solving

$$|S|_{t+1} = |S|_t \; |T|$$

and plotting the value of S_3 at each iteration.

Experimental Plan

The simplified Markov process model was evaluated under two types of stimulus conditions using eye movement measures. Abstract scenes provided highly-controlled features for all objects in the scene and eliminated context information. Realistic scenes provided more complex and variable features for all objects, as would occur under natural conditions.

The focus of the initial tests was on the validity of the underlying assumptions about object features, processing components, and expectations. There was a direct correspondence between the types of object features in the Abstract scene experiment and the Realistic scene experiment. The objects used in the Abstract scene experiment were based on the features extracted from actual tactical target images, and it was expected that the effects of those features on the two major processing components would generalize to the realistic conditions. In the following sections, the two experiments are presented, and the preliminary results as well as detailed plans for a complete model evaluation and analysis are presented.

ABSTRACT SCENE EXPERIMENT

Introduction

The abstract scene experiment examined the basic feature processing components of the simplified two-stage Markov process model under controlled conditions, while maintaining maximum generalizability to target search and detection under realistic conditions. Manipulation of input data and observer expectation provided a means of assessing the impact of object feature perception on the target acquisition process. The stimuli were constructed to maintain many of the important characteristics of real targets and clutter, thereby making it possible for the results to contribute to an understanding of the perception of real scenes.

The experiment manipulated two of the four aspects of input data, target and clutter, as well as observer expectation. Five characteristics of target and clutter were systematically varied: size, contrast, orientation, contour, and internal brightness modulation. The first three were global object characteristics of relatively low spatial frequency content and were expected to be processed perceptually as low-level features. Contour and internal brightness modulation, on the other hand, contained higher spatial frequency data and could be expected to elicit a relatively detailed examination for their perception, thereby qualifying them as high-level features. As will be discussed later, the target could always be determined on the basis of its specific high-level features.

The degree of pre-trial knowledge concerning the low-level features of the target to be detected provided two levels of operator expectation. As presented earlier, expectation is the sum of general life experience, specific training, and a priori knowledge concerning the task being performed. In the present experiment, the nature of the abstract scenes minimized the impact of prior life experience resulting in observer expectation being primarily determined by the briefing information given prior to each experimental trial.

Context and texture were excluded as input data sources from the abstract scenes to allow a critical examination of the effects of target and clutter characteristics and observer expectation. Because context can have a major influence on the general search strategy developed during the initial orientation stage of perceptual processing, its inclusion in the present experiment would have, at best, made the assessment of target and clutter feature effects considerably more difficult. Further, the addition of context would have increased the relevance of prior experience which would have been confounded with the experimental manipulation of target briefing specificity. Finally, context is more appropriately investigated in a separate experiment which builds upon the results of the present investigations. Texture was also not considered in the present study because of its relative unimportance compared to the other aspects of input data.

The data from this experiment were intended to allow examination of four major questions related to the processing components of the two-stage Markov model. First, does the model adequately describe the obtained data when the transition probabilities are determined on the basis of the sequence, pattern, and duration of eye fixations on individual objects? Second, is there evidence for a hierarchy of feature processing when searching for the target? Third, does the perceptual processing change when the low-level features of the target are known a priori? Fourth, do high-level features elicit detailed examination? The latter three questions were examined using the number and duration of fixations on objects as a function of their feature characteristics.

Method

The eye positions of 10 observers were recorded as they searched 27 abstract scenes each containing a specific target and 27 clutter objects. Target objects had one of nine combinations of low-level features and a single set of high-level features. Each clutter object was one of a possible 81 consisting of 27 combinations of low-level features and the three sets of non-target high-level features. The 10 subjects formed two groups of five subjects each. One group received a general briefing which described the high-level features of the target. The second group received a specific briefing prior to each trial which added information regarding the low-level features of the target.

Experimental Design. A mixed-factor factorial design was employed to assess the effects of two briefing conditions, nine targets, and 81 clutter feature combinations. The factorial combination of three sizes, three contrasts, three orientations, and three sets of high-level features resulted in the 81 possible clutter objects.

Ideally, all 27 combinations of target low-level features would have occurred with each possible clutter object resulting in 27 images with one target and 81 clutter objects each. The number of images would have been acceptable; however, a total of 82 objects on a single image would have resulted in an unacceptably high density. The number of objects per image was reduced by blocking the clutter objects into three sets of 27 each. The objects in each block were determined using a partially balanced factorial design (Kirk, 1968, 1. 356)[40] with the third- and fourth-order statistical interactions confounded with block.

Blocking in this manner allowed all comparison between target and clutter conditions to be made, but a complete replicate required three images per target. Examination of all 27 targets with three images per target would have required a total of 81 trials and three experimental sessions per subject. A further reduction in the design was made by reducing the total number of targets to nine. The selected targets were the eight combinations of extreme values of size, contrast, and orientation plus the mid-values of these three low-level feature variables.

The final design had two groups of five subjects; each group viewed a total of 27 images with one of nine possible targets and 27 clutter objects. Over the 27 images, each subject observed every possible clutter object with all nine possible targets. Briefing was a between-subjects variable with one group receiving specific target information and the other group receiving only general target information.

Stimuli. The target and clutter objects were systematically abstracted from photographs of actual vehicles to retain as many realistic characteristics as possible. The original targets were broadside views of three vehicle types - tank, truck, and APC - with an aspect consistent with a 910m altitude and sensor depression angle of 0.35 rad. The steps in the abstraction process are shown in Figure 6 and detailed below.

A rectangle was circumscribed around the original photograph of the vehicle. The height of the rectangle was the distance from the longest flat surface on the upper portion of the vehicle to the bottom of the tire or tread surface. The width of the rectangle was the distance between the extremes of the main body of the vehicle. With the exception of the tank gun, all three vehicles have approximately the same four-to-seven height-to-width ratio. The circumscribed rectangle was converted to a polygon by retaining those portions of the rectangle which contacted the vehicle image and connecting the ends of these line segments with additional straight lines. The straight lines were then replaced by wavy lines approximating a sine wave with an amplitude of 5 percent of the object width and with about three cycles per object width. This procedure was designed to simulate a degradation of the target similar to that obtained by Williams et al.[73]

The basic outlines were next modified to achieve two internal brightness modulation high-level feature conditions. Either a uniform white bar with a height-to-width ratio of five-to-nine or a checkerboard pattern was added to the center of the stimulus image. The checkerboard pattern was composed of white squares with sides equal to 7 percent of the width of the stimulus object. The 20 squares of the checkerboard pattern had an area

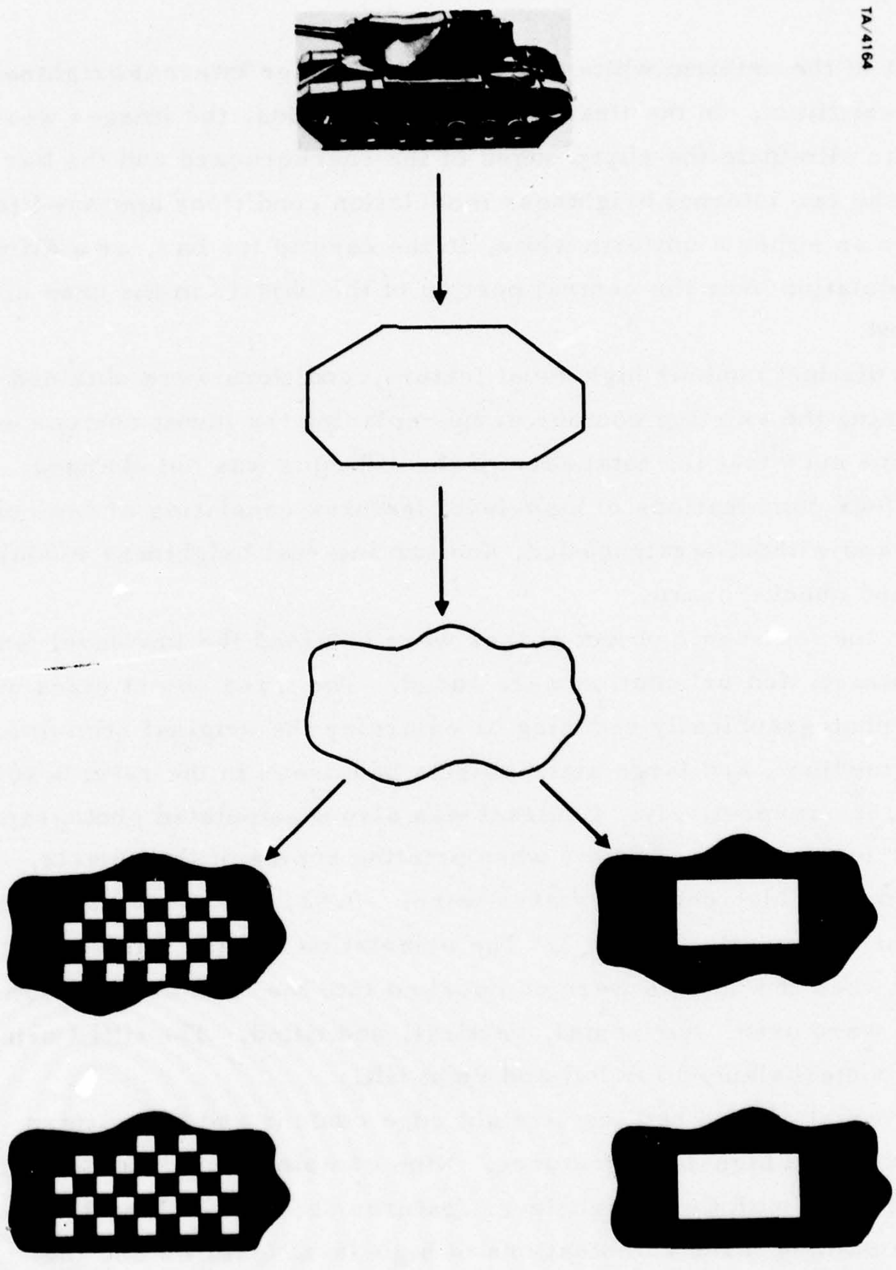


Figure 6. Steps in the generation of the abstract objects.

equal to that of the uniform white bar used in the other internal brightness modulation condition. In the final stimulus preparation, the images were defocussed to eliminate the sharp edges of the checkerboard and the bar patterns. The two internal brightness modulation conditions appeared to the observer as either a uniform white, in the case of the bar, or a diffuse internal modulation over the central portion of the object, in the case of the checkerboard.

Two distinct contour high-level feature conditions were obtained by either retaining the existing contour or by replacing the lower contour with a straight line such that the total area of the stimulus was not changed. Thus, there were four combinations of high-level features consisting of two contours, with and without straight edge, and two internal brightness modulations, bar and checkerboard.

Once the four basic object shapes were obtained the low-level features of size, contrast, and orientation were added. The three object sizes were obtained by photographically reducing or enlarging the original stimulus. The small, medium, and large sized objects had areas in the ratio 0.46, 1.00, and 1.46, respectively. Contrast was also manipulated photographically by varying the exposure time when printing copies of the objects. The low, medium, and high contrast values were: -0.92, -4.2, and -5.4 using the definition $-(B_{max}-B_{min}/B_{min})$. The orientation of each stimulus was manipulated when the objects were composited into the final arrays. Three orientations were used: horizontal, vertical, and tilted. The tilted orientation was counterbalanced for left and right tilt.

The target always had the straight edge contour and bar pattern internal modulation high-level features. Nine combinations of low-level features combined with these high-level features defined the target object set. The remaining three combinations of high-level features and the 27 combinations of low-level features defined the clutter object set.

The stimulus arrays each consisted of 27 clutter objects and one target as shown in Figure 7. The location of objects within the array was

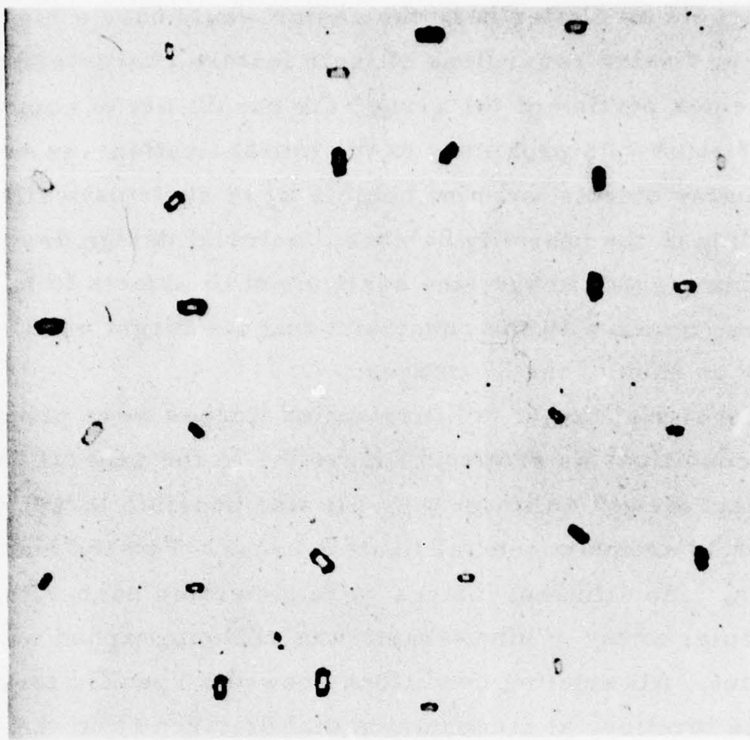


Figure 7. Typical abstract scene.

the same for all images. The locations were determined by dividing the display into a six-by-six matrix of equal square cells. The 36 cells were reduced to 28 by eliminating the corners and four center cells from the matrix. To avoid having systematic rows and columns of objects, the locations were determined by randomly offsetting the center of the 28 cells one medium sized target width in one of eight radial positions defined by the secondary points of the compass.

Stimulus objects were eliminated from the center for two reasons. Meta-contrast, or forward masking of the objects, might result from the presence of the center fixation cross used to stabilize the eye movements in the pre-trial period.[25] Second, because subjects started at the center

of the display, targets or clutter near the center would have a higher probability of being fixated regardless of their feature characteristics. By eliminating the center portion of the array, the possibility of confounding feature characteristics with proximity to the initial fixation was reduced.

The 81 clutter objects and nine targets were systematically assigned to images according to the partially balanced factorial design described previously. Within a given image, the assignment of objects to locations was randomly determined with the constraint that the target occur in a different location on each of the 27 images.

Separate pre-trial target familiarization images were prepared for the two briefing conditions as shown in Figure 8. In the general briefing condition, a subject viewed an image with all nine possible target objects arranged in a circle around a central fixation cross. For the specific briefing condition, nine different images were generated each with one target. The circular array of nine targets was rephotographed with eight targets masked out. All briefing conditions showed a specific target in the same absolute location, at a constant radial distance from the starting fixation cross.

When viewed by the subject, the total stimulus arrays subtended a visual angle of 250 mrad. The medium size objects were 5.0 mrad high by 8.8 mrad wide which is the same as the realistic targets used by Scanlan (1976).[54] The large objects were 6.0 by 10.6 mrad and the small objects 3.4 by 5.9 mrad. The blank center portion of the display had a radius of approximately 44 mrad, and the distance between the centers of individual objects was at least 30 mrad.

A calibration image was also prepared which consisted of a uniform field with a five-by-five square matrix of small (2 mrad) black dots equally spaced and covering an area 16/15 the total area of the experimental image. The calibration matrix was made larger than the experimental images to insure measurement of all extremes of the image during actual testing.

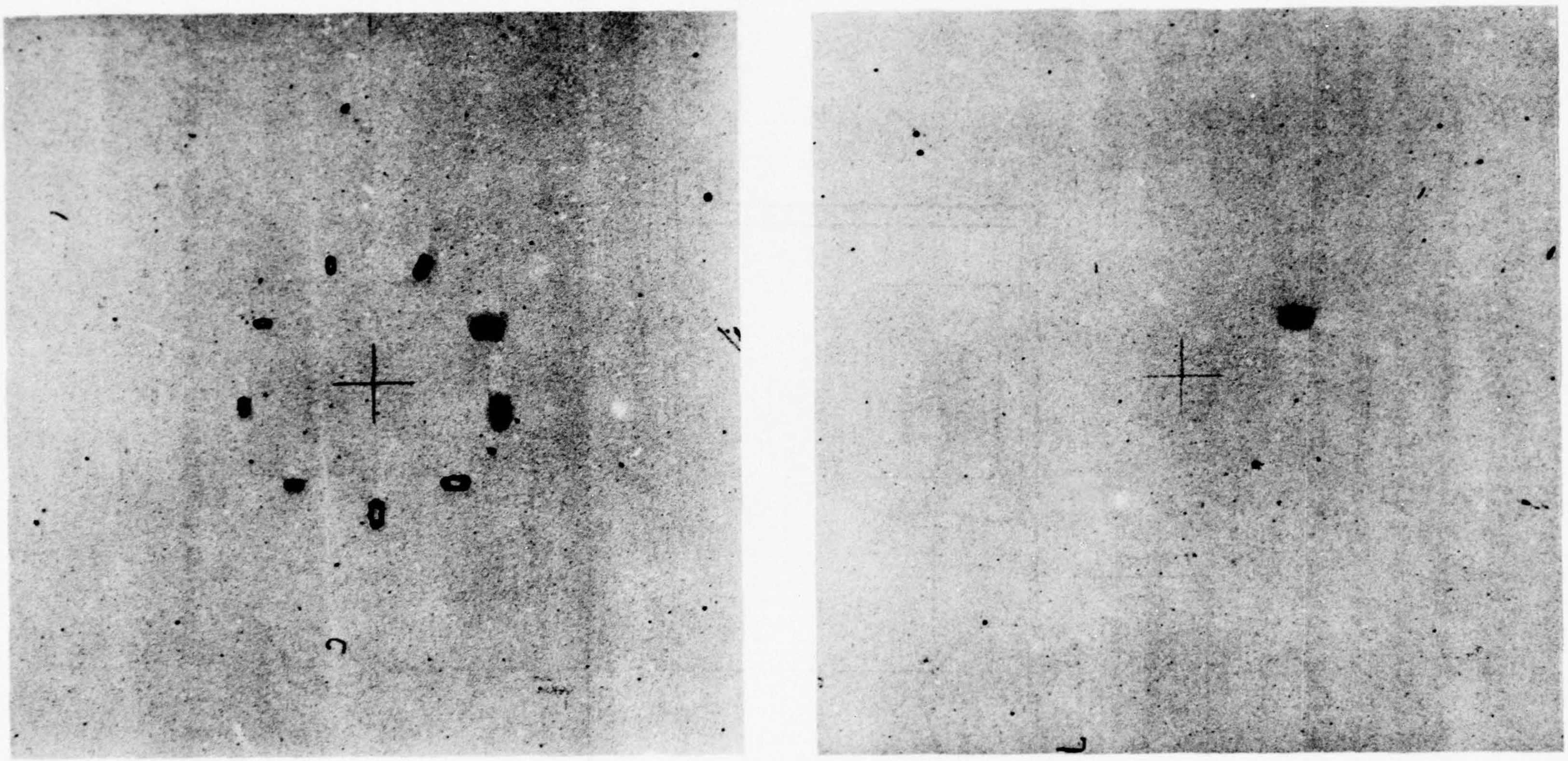


Figure 8. Typical pre-trial briefing and fixation cross images.

Apparatus. The experimental apparatus is shown schematically in Figure 9. The major elements were the Stanford Research Institute Perkinji Image eye tracker with associated controls, a two-field projection tachistoscope, and an Ampex PR-1300 FM magnetic tape recorder.

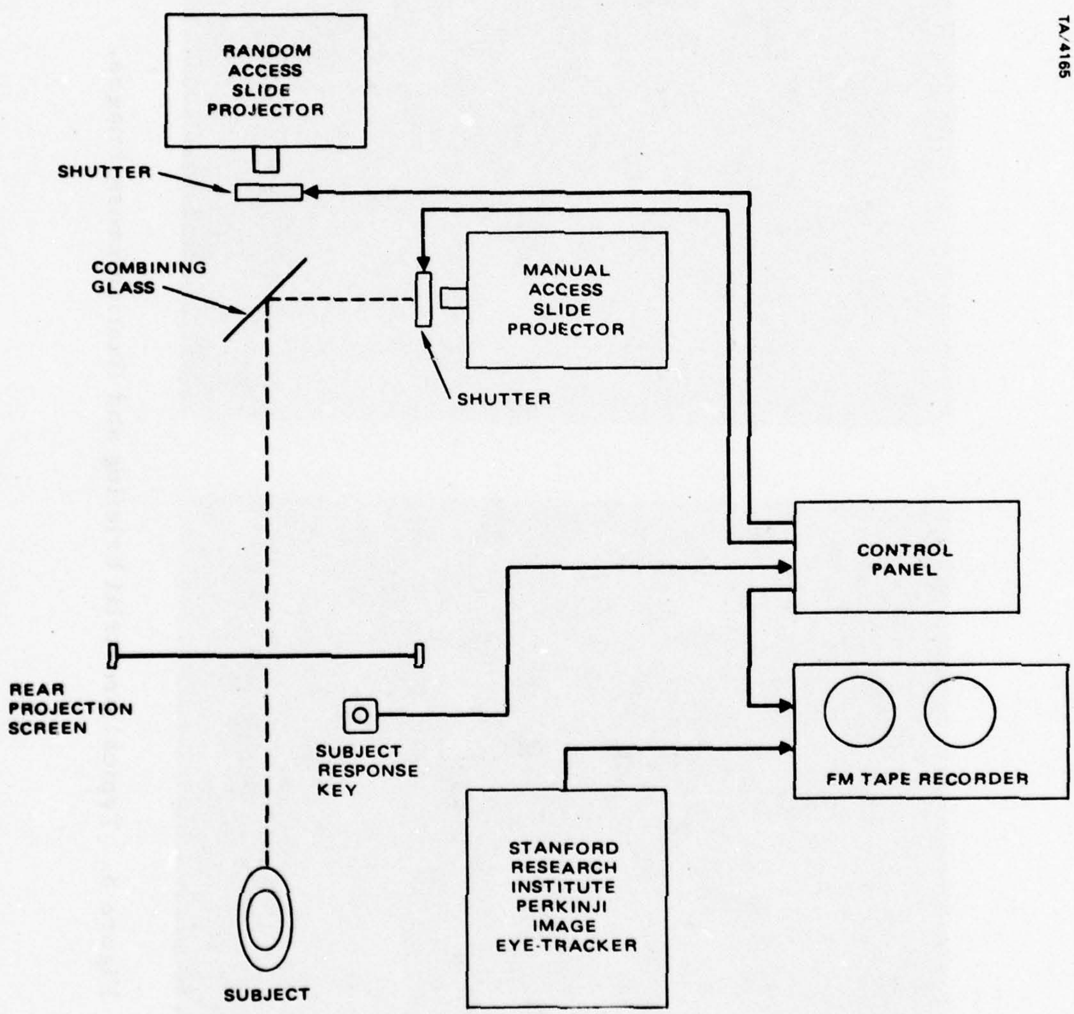


Figure 9. Schematic diagram of experimental apparatus.

The subject was seated behind a special bench which supported the eye tracker and an adjustable bite bar mount to provide head stabilization. The horizontal and vertical eye position and blink signals were recorded by the magnetic tape recorder. The eye position signals were also connected to the horizontal and vertical deflection inputs of an oscilloscope to allow the experimenter to monitor a subject during a trial.

The tachistoscope consisted of two Kodak Carousal 35 mm slide projectors, two Gerbrands model G1166 shutters, a combining glass, and a rear projection screen. A random access slide projector was used in the stimulus field to allow rapid selection of the desired image. A conventional manual select side projector was used for the briefing and fixation field. Shutter control was provided by a special control box which also generated signals, recorded on the magnetic tape, indicating the trial number and the duration of a trial.

The subject viewed the stimuli on the rear projection screen at the distance of 1 m. The overall luminance of the two fields were matched at 10.6 cd/m^2 using neutral density filters. Ambient illumination was a constant 58 lx for all subjects.

<u>Subjects.</u> Ten adult subjects, four male and six female, were recruited from the Stanford Research Institute staff and the Stanford University student population. All subjects were tested for uncorrected 20/20 visual acuity. They were paid for their participation.

<u>Procedure.</u> Each subject participated in one session lasting approximately one hour. Upon arrival, the subjects were tested with a standard Snellen chart to confirm a normal visual acuity rating of 20/20 or better. The session then proceeded with a briefing session of about 10 minutes.

A subject was given a standard information sheet, specific instruction sheet, and an informed consent form to sign. The standard information sheet explained the general purpose of the research and the nature of the eye tracking apparatus. The consent form was presented and the subject's questions were answered so that fully informed consent to participate could

be given. The specific instructions which explained the experimental task were presented last.

The procedures for apparatus adjustment and calibrations included fitting the bite bar, a metal form coated with dental impression compound, and adjustment of the head support. The adjustment was made for maximum subject comfort. The subjects were always allowed to rest and release the bite bar between trials if they felt fatigue. The optical and electrical alignment of the eye tracking device was accomplished in about 10-15 minutes after the fitting of the bite bar and headrest.

Eye tracker calibration data were collected both before and after the experimental session by having the subject fixate each of the 25 dots of the calibration slide. This procedure provided the data necessary to compensate for slight non-linearities in the Perkinji Image tracker and for any differences among individual eyes.

Ten training trials proceeded the experimental trials. For each trial a briefing slide was initially presented. In the general briefing condition this slide contained all nine possible targets. In the specific briefing condition, only the target for that trial was shown. This visual briefing was supplemented with a verbal description of the target to be detected. The subject searched for the target and pressed a switch when he had detected it. Feedback on the correctness of the response was given during training.

The 27 experimental stimuli were presented in randomized order without knowledge of results. The experimenter monitored eye movements on the oscilloscope and did not start a trial until the subject steadily fixated the center fixation cross. The image was on until the subject made a response and automatically turned off through shutter control.

Upon completion of the experimental session, a subject was de-briefed and allowed to ask questions. No rating of a subject's performance was given. The information released to the subjects was of a general nature; none of the subjects tested requested detailed or sensitive information.

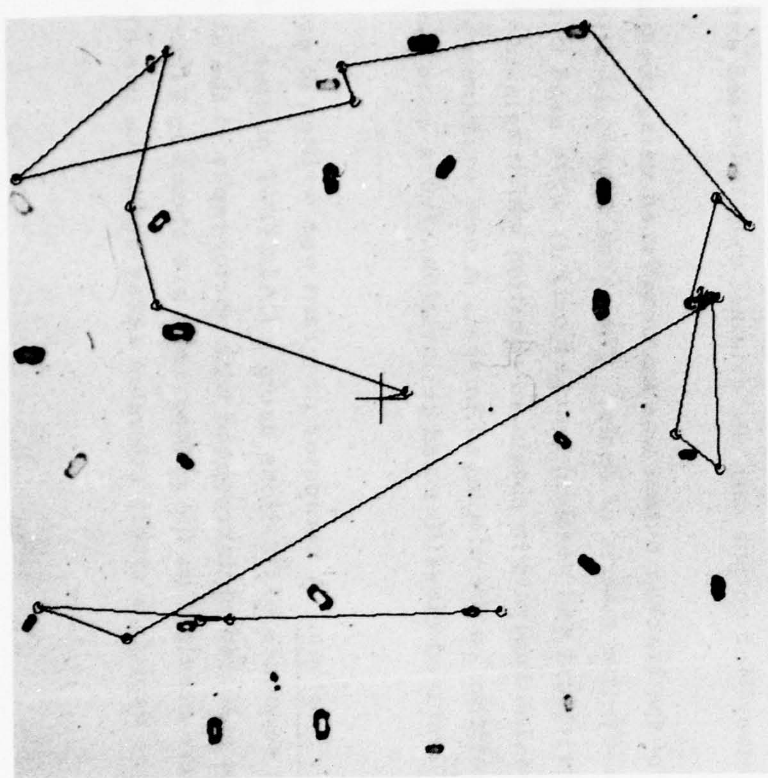
Results and Discussion

Data Reduction. The raw eye position data on the analog magnetic tape were subjected to several data processing steps to obtain the final sequential fixation position and duration data for each trial. The seven channels of analog data were converted to digital form at a sampling frequency of 200 Hz. The digital tapes were used as input to a pre-processing program which eliminated data between trials, extracted the trial number, and saved the low-pass filtered horizontal and vertical eye position data in a disc file. This disc file was then used as input to a classification program which determined when the eye was fixating and output the average fixation position and its start and end time relative to the beginning of a trial. These data were stored in another disc file keyed by subject and trial number.

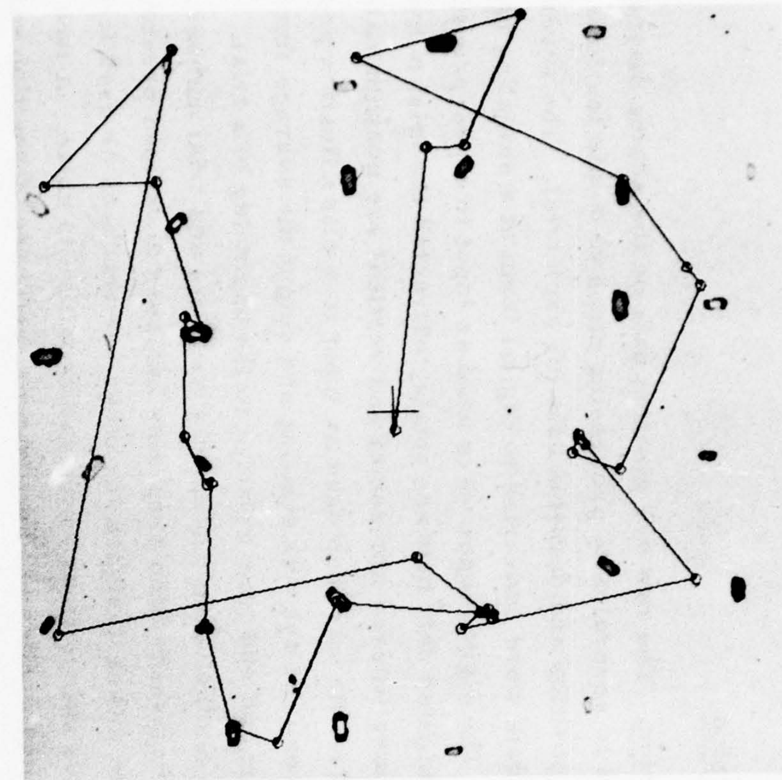
A number of classification programs had been developed previously, however, these were either designed for other eye trackers or used first and second derivative calculations which are sensitive to noise. A new classifier was developed which used a five point estimate of position variance from the current fixation to determine if the eye was stationary or moving. Verification of the classifier output with the original data indicated excellent correspondence.

Calibration of the tracker output was accomplished using the subject fixations on the five-by-five matrix of spots. The actual known location of the spots and the horizontal and vertical tracker outputs were used in a second order regression analysis to obtain an equation which related tracker output to absolute position on the displayed image. A new equation was fit for each subject. Combined classifier and calibration errors were less than ±3.5 mrad.

Sequence of Fixations. A computer program was written to graphically present the position sequence of fixations using a CALCOMP plotter. The scale of the plot was selected to correspond with photographs of the stimulus arrays. Typical plots overlaid on the photographs are shown in Figures 10 to 13. Each figure depicts a single stimulus array under the two briefing

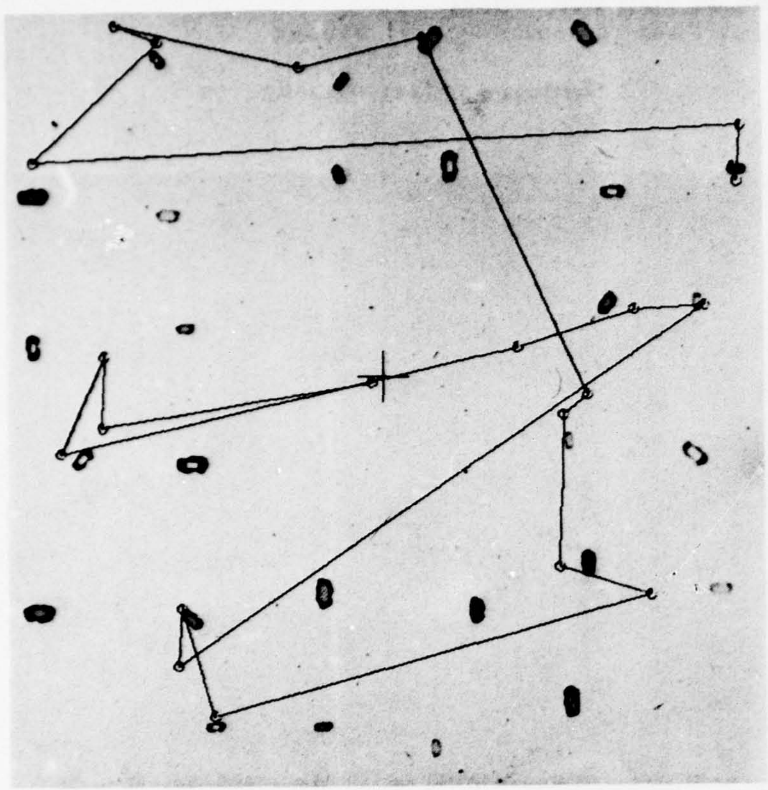


a. Specific target briefing.

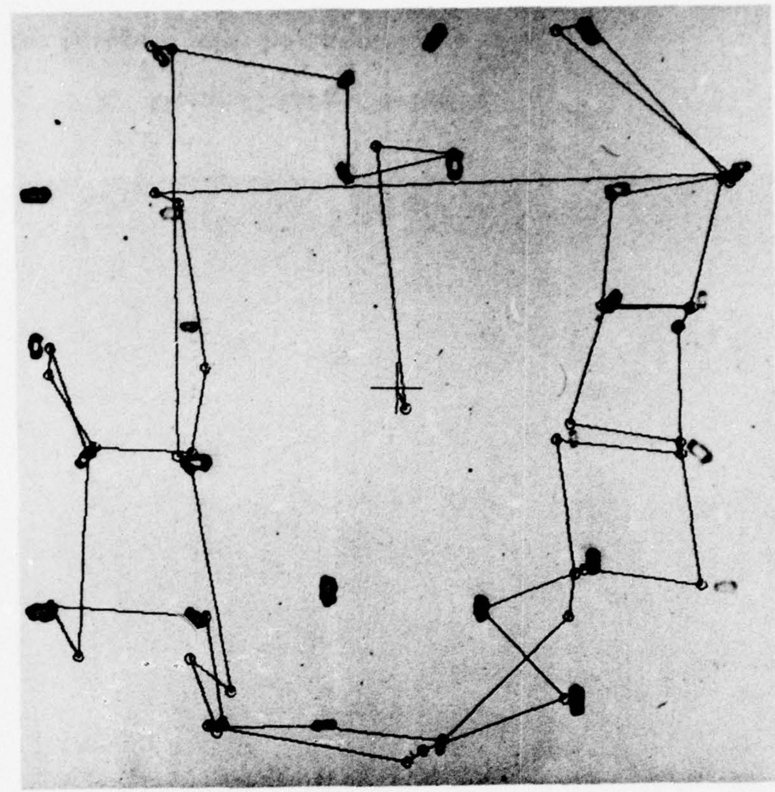


b. General target briefing.

Figure 10. Sequence of eye fixations for Image 2 with horizontal, small, and light target.

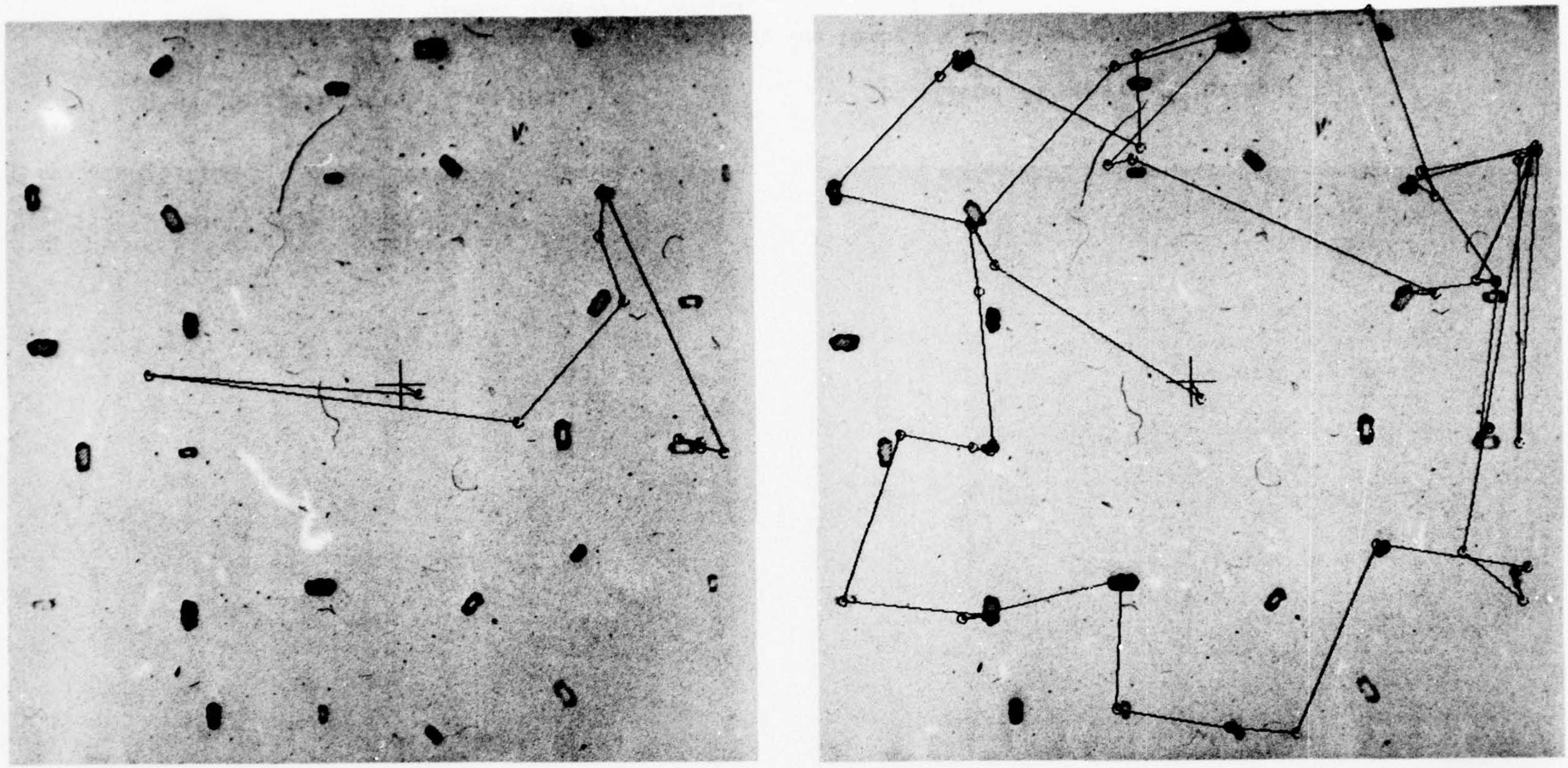


a. Specific target briefing.



b. General target briefing.

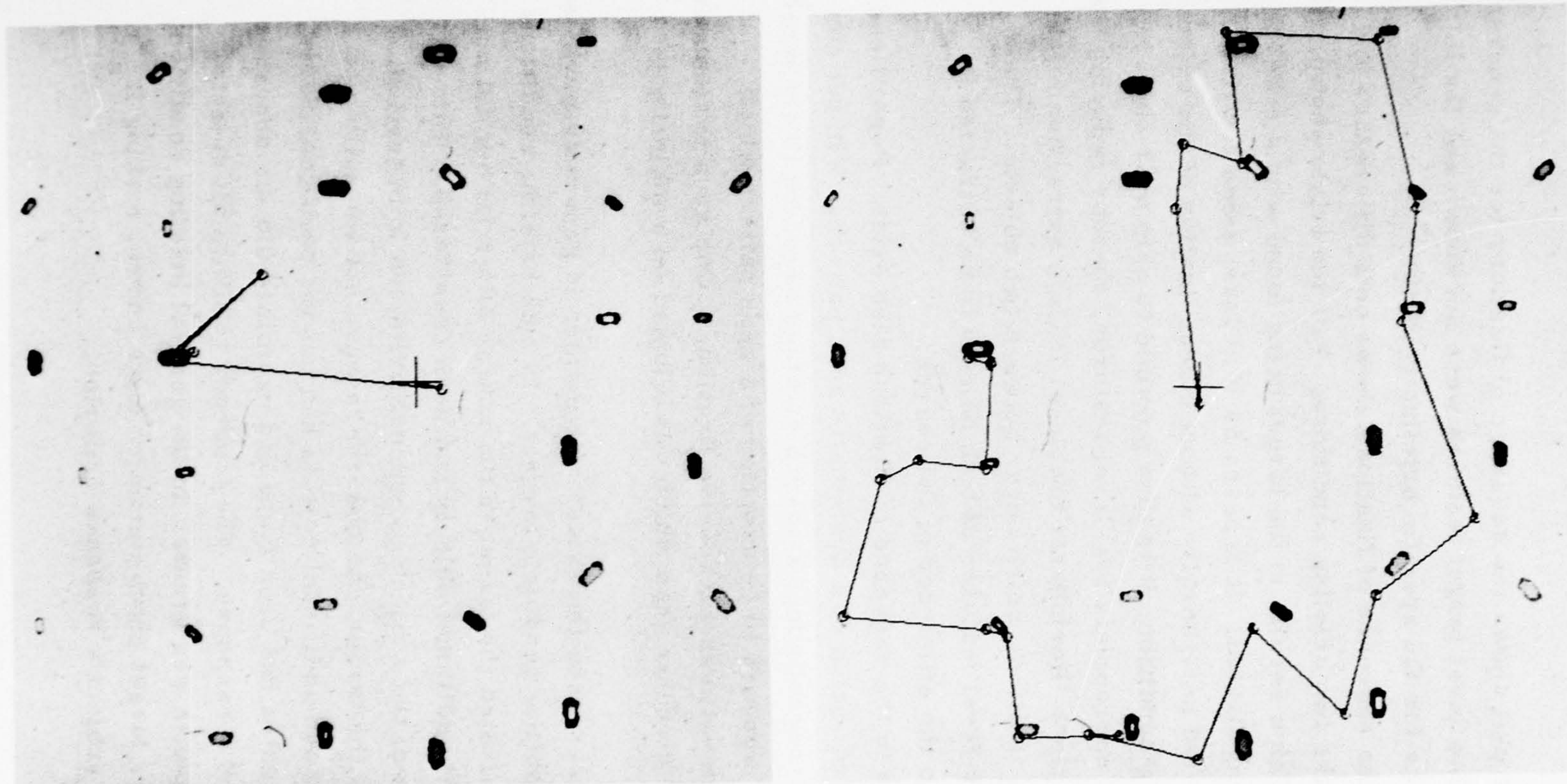
Figure 11. Sequence of eye fixations for Image 4 with horizontal, small, and dark target.



a. Specific target briefing.

b. General target briefing.

Figure 12. Sequence of eye fixations for Image 6 with horizontal, large, and light target.



a. Specific target briefing. b. General target briefing.

Figure 13. Sequence of eye fixations for Image 8 with horizontal, large, and dark target.

conditions. The right plot shows the sequence of fixations for the general briefing in which the low level target features were not known and the left plot shows the sequence for the specific briefing condition.

Inspection of the sequence of fixations shows both differences and similarities between the two briefing conditions. For the trials shown, the specific briefing condition results in the target being found with a fewer number of objects being fixated. It appears as if at least some low-level features can be extracted peripherally without direct fixation of the object. In the specific briefing condition, it is thus possible to skip over objects which do not have the appropriate low-level features, thereby reducing the number of objects fixated. Peripheral low-level feature extraction might also account for those fixations which occur between two objects. These fixations allow the low-level features of both objects to be extracted in parallel and can add to the efficiency of the search.

Commonalities in the sequence of fixations also exist. Regardless of briefing, there is an indication of a systematic scan pattern. Without context cues the observer cannot adopt a search strategy based on likely target locations. The result appears to be adoption of a scan pattern which minimizes the distance between successive fixations. One such systematic pattern is a clockwise circular scan which was adopted on a majority of trials.

As the trial progresses the systematic pattern is generally modified to re-fixate high probability candidate targets. In both briefing conditions the subject generally fixated the target in the course of a trial but did not choose to respond until additional objects had been considered. This is, of course, an example of the subject's response criterion in operation.[66] On the first fixation of the target, the observers were not yet willing to make a response, because additional objects had not yet been considered. Only after other candidates had been found and examined did the observer make a final choice and a response. The tendency to fixate all objects prior to making a response was greater in the general briefing condition probably because fewer target characteristics were known, making it difficult to exceed the subject's response criterion.

Knowledge concerning the multiple features on which to base a detection decision not only reduced the time spent detecting the target but also improved the probability of being correct. Figure 14 shows the cumulative probability of correct detection as a function of time for the two briefing conditions. The specific briefing condition dramatically improved target detection performance. The probability of correct detection improved from 0.6 to 0.85 and the time for a 0.5 probability of detection dropped from 18 sec to 6.4 sec with the specific briefing.

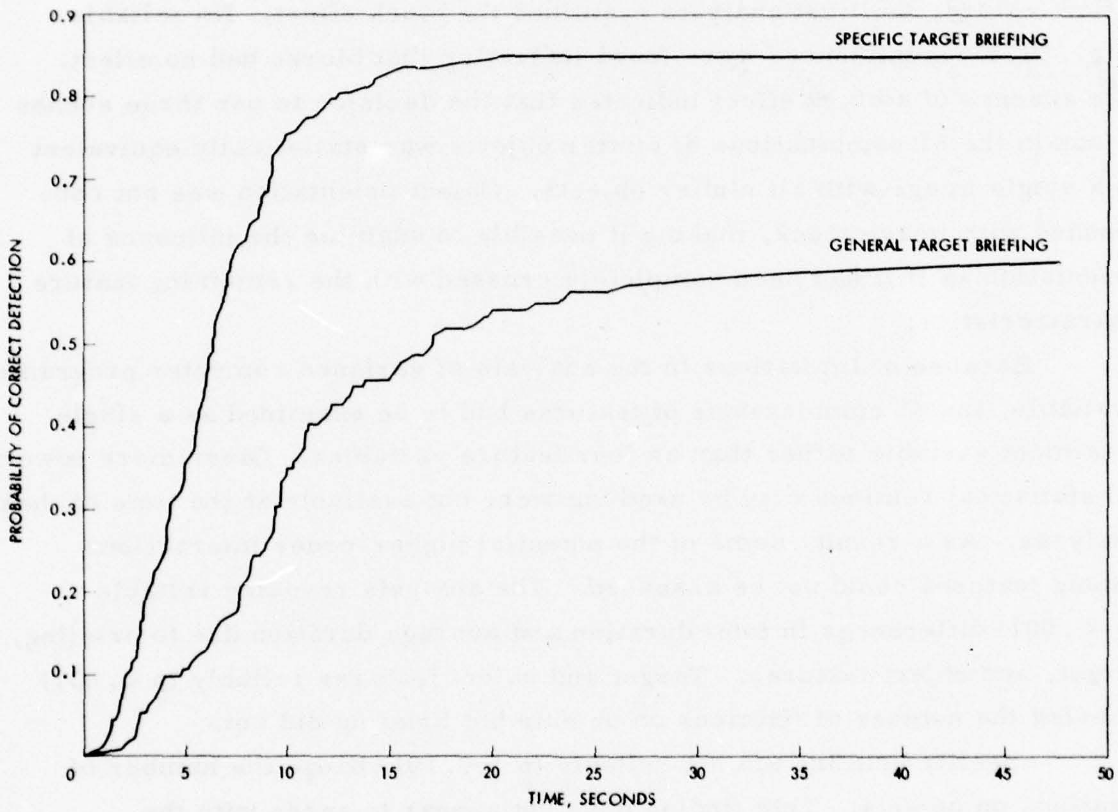


Figure 14. The effect of briefing on the probability of correct detection as a function of time.

Analysis of Variance. Analyses of variance were performed on the number and duration of fixations on clutter objects. The data for these analyses were obtained by classifying the object being fixated on the basis of the known object locations. A 30 mrad diameter area about each object was defined, and any fixation within that area was assigned to that object. The number of fixations, total duration of fixations, and the average fixation duration, were calculated as a function of briefing, target, stimulus block, and feature characteristics of the object. Eighteen percent of the fixations were eliminated as not being within 30 mrad of an object.

Because the stimulus block was partially confounded with the orientation feature, the first analysis examined the block effect. No reliable ($p \geq .25$) block influences were found indicating that blocks had no effect. The absence of a block effect indicates that the decision to use three scenes to obtain the 81 combinations of clutter objects was statistically equivalent to a single image with all clutter objects. Object orientation was not confounded with image block, making it possible to examine the influence of orientation as if it had been completely crossed with the remaining feature characteristics.

Because of limitations in the analysis of variance computer programs available, the 81 combinations of features had to be examined as a single treatment variable rather than as four feature variables. Other more powerful statistical routines may be used but were not available at the time of these analyses. As a result, some of the potential higher-order interactions among features could not be assessed. The analysis revealed reliable ($p < .001$) differences in total duration and average duration due to briefing, target, and object features. Target and object features reliably ($p \geq .001$) affected the number of fixations on objects but briefing did not.

Specific briefing did not reliably ($p > 0.10$) change the number of fixations on objects. This finding does not appear to agree with the observations made previously based on an examination of the sequence of fixations presented in Figures 10 to 13. The most likely explanation for the

discrepancy lies in the manner used to obtain the data for the analysis of variance. Each time an object was fixated the number of fixations for that object was incremented by one. The number of fixations used in the analysis of variance, therefore, does not correspond to the number of objects fixated. As calculated, it is not possible to distinguish a large number of fixations on a few objects from a single fixation on many objects. Further analysis will be needed to rectify this ambiguity.

Specific briefing reduced the total time spent looking at clutter objects as well as the time per fixation on clutter objects. When the low-level features of a target are known, it is possible for the observer to reject clutter objects more quickly than can be accomplished when only high-level features are known. The low-level features can be extracted in less time than the high-level features supporting the basic hypothesis of levels of feature analysis. The concept of feature levels was critical to the formulation of the target search and detection model previously presented, and these results lend definite support to the hypothesis that low-level features impact search and high-level features impact examination. These data also demonstrate the influence expectation can have on the processing of input data. Recall that the scenes used in this study had no context cues which reduces the impact of the prior experience aspect of expectation. With realistic scenes, context will be present and expectation will likely have an even larger influence on search and feature extraction.

The characteristics of the object being examined also had a significant effect on the time spent looking at the object. Some objects had an average duration of 217 msec per fixation while others required only 66 msec per fixation. Overall, some objects were considerably easier to accept or reject as targets than were other objects. The feature characteristics producing these differences have not yet been examined in any detail; however, it is likely that both size and contrast are contributors. The effect indicates that differences in ease of extraction exist within the high and low-level categories of features.

The target to be detected made a difference in performance also. The exact nature of this effect cannot be determined from the present analysis. It is likely, however, that large, high-contrast targets were detected more easily because of the higher visibility of the high-level features. Because feature information could be extracted more easily for these targets, it was possible for the observer's response criterion to be reached either sooner or with fewer repeated fixations. Trials where the high-level features of the target were more visible would have been shorter in length thus reducing the number of possible fixations. It is also likely that the increased visibility of high-level features reduced the time needed for their extraction, thereby producing changes in the duration of fixations on objects. Examination of these issues will require further analyses of the data using dependent variables which are less ambiguous with respect to these effects.

Although the preliminary analyses performed leave a number of questions unanswered and do not even begin to exploit the data, they do provide initial support for the major hypothesis underlying the proposed model. The factors hypothesized to have an effect on the number and duration of fixations on clutter objects were all found to be reliable. Further the results are in the expected direction and the magnitudes of the differences are large.

The data clearly support the hypothesis that target search and detection is influenced by both the feature characteristics of individual objects and by the expectation of the observer. When the specific low-level features of the target are known, the duration of fixations on each object are reduced, because most objects may be rejected as not being the target on the basis of these features. When only a general briefing is given, the durations are longer because the high-level features must be extracted before an object can be rejected as the target.

The general hypotheses of the multi-component model of target search and detection are supported by the analyses conducted. However, more detailed analyses must be undertaken to fully exploit the obtained data.

An analysis of the specific feature characteristics and their effect on eye fixation probability and duration is an obvious first step. Further analyses which consider the similarity of clutter to targets and the impact of location and features on the sequence of fixations as well as the duration and probability of fixation are needed. It is also recommended that the probability of next fixating an object based on the characteristics of that object and the current eye position be examined. Analyses such as this can be used to understand the specifics of the feature extraction process and should provide a means for describing the differences between those trials which required an extensive number of fixations and those which required only a few fixations.

In addition to analyses which examine the feature extraction and expectation processes, it is possible to use the eye fixation data to estimate the transition probabilities in the Markov model of Figure 5. Suitable, operational definitions for the various states must be obtained and applied to the data to allow unambiguous identification of the search and examination states. It will then be necessary to verify that the data meet the underlying assumptions of a Markov process and, finally, that the estimates of the transition probabilities yield a predicted cumulative probability that is the same as the obtained curves shown in Figure 14. Initial, informal examination of the data suggests that the Markov model can be expected to yield good results.

REALISTIC SCENE EXPERIMENT

Introduction

The Realistic Scene experiment was designed to provide data to evaluate the predictive power of the two-component Markov process model under realistic target and background conditions, to determine the generality of results obtained in the Abstract Scene experiment, and to suggest the important feature attributes of real scenes. The experiment yielded performance data and eye fixation measures which could be analyzed in terms of the Markov process model. Key questions to be explored were:

- Is there evidence for search and examination processing components under realistic conditions, and is the evidence consistent with that obtained under the Abstract Scene study?
- What is the effect of context, clutter and target features on the Markov process model parameters?
- Is the two-component Markov process model an adequate predictor of observed performance under realistic conditions?
- Which aspects of realistic scenes qualify as perceptual information and what are their feature attributes?

The Realistic Scene images contained all aspects of input data. The inclusion of context information in the task was an important extension of the conditions previously examined in the Abstract Scene experiment. Context was expected to have a significant effect on the search state parameters. The realistic target and clutter object features were also an extension of conditions to those likely to be encountered in real world applications. High-level features in target and clutter were expected to have relatively large effects on the examination state parameters.

The experiment included a wide range of target and background conditions. The overall complexity of the background, the position of the target within the scene, and target contrast were varied. Target size was held constant, and briefing information was constant for all subjects. Because the background scenes were photographs of natural terrain, certain background characteristics were not experimentally manipulated, but were measured through a multiple-rater procedure. The ratings, similar to the scene metrics developed in a previous research effort,[54] identified sub-areas within each background scene which contained various categories of perceptual information such as roads, fields, and trees. The input data in the task, therefore, varied across experimental trials, and it was possible to examine the effect of type of information on the two-component Markov model parameters.

Method

Stimuli. The stimuli consisted of tactical vehicle target images embedded in aerial black and white photographs of rural New York State, simulating Central European terrain. The background scenes were selected from images employed in a previous experiment.[54] The scenes were photographed at an elevation of 910m and mean camera depression angle of 0.35 rad. A field of view of 140 mrad x 140 mrad was employed.

The targets were HO scale models of tactical vehicles: M-60 tank, 2.5-ton truck, and armored personnel carrier (APC). The models were photographed against a uniform white background. The internal modulation of each target was artificially enhanced by an artist, and appropriate shadows were added. Compositing the selected target and background images was accomplished by superpositioning the target over the background and re-photographing to 35mm format. Typical images are shown in Figure 15.

Figure 15. Typical experimental scenes.

Two sets of 10 background scenes were prepared with the target located in different positions for each set. One set of locations was identical to those used in a previous study[54] while the other set was selected to systematically examine the effect of target location within the scene. Two backgrounds without targets and 10 training images were also prepared.

The stimuli were presented at the same 250 mrad angular subtense and 1 m viewing distance used in the Abstract Scene experiment. The targets subtended a visual angle of 5.0 by 8.8 mrad which is the same as the medium-sized object in the Abstract Scene study.

A calibration image was prepared which consisted of a uniform field with a 5 x 5 matrix of black dots (1.5 mrad) equally-spaced and covering an area 16/15 the total area of the experimental image display. The calibration matrix was made larger than the experimental images to insure measurement of all extremes of the image during actual testing.

A fixation cross (30 mrad) on a uniform background was prepared to coincide with the center of the experimental image display. The fixation cross was presented before each trial so that the subject would be fixating the exact center of the display at the beginning of each search task.

Apparatus. The apparatus for the Realistic Scene experiment was identical to that employed in the Abstract Scene experiment. The display size was identical, and the overall luminance of the display was adjusted with Wratten filters to match the level used in the previous experiment ($10.6\ cd/m^2$). Ambient luminance levels were also held constant at 58 lx.

Subjects. Twenty adult volunteer subjects were obtained from the Stanford Research Institute staff and the Stanford University student population. All subjects were tested for a 20/20 visual acuity rating. All subjects were paid for participation. There were six male and 14 female subjects 10 of which also participated in the Abstract Scene experiment.

Experimental Design. A balanced experimental design was not employed, due to the difficulty of controlling all variables of interest within a set of realistic background scenes. A set of targets and backgrounds were selected which would represent a range of values for the variables of interest. The effects of varying information within the task on the processing components was of primary interest in the experiment, and they could be assessed from the eye fixation data without a completely crossed design.

The experimental images included 10 background scenes of varying levels of subjectively rated complexity, determined for that set of images in a previous study.[54] Each background was prepared with a single embedded target in a realistic location with respect to context, scale, and shadow. Then, the 10 images were prepared with the same target moved to a different location within the scene, and in a different type of context. If the first target location had been in a likely area, based on terrain context, it was moved to an unlikely area and vice versa. The likelihood of fixating an area as a *function of context* was determined by type of context features and terrain pattern.

The two sets of images were presented to different groups of 10 subjects each. Two catch trials were also presented, which consisted of scenes of moderate complexity without tactical targets embedded in them.

All subjects received randomized presentations of 10 experimental images and 2 catch trials. A Latin Square randomized procedure was used to determine the order of presentation of the 10 experimental images.[17] The catch trials were inserted into the randomized orders in a counter-balanced scheme such that each catch trial appeared with equal frequency over all trials and over the first-half of the experimental session.

Procedure. The experimental session replicated the procedures employed in the Abstract Scene experiment. All subjects were initially tested for a visual acuity rating of 20/20 with a standard Snellen chart, and were given information sheets containing general information, informed consent forms, and subject instructions specific to the experiment. The

procedures for apparatus adjustment and calibration were identical to those used in the Abstract Scene experiment.

The experimental session began with five training trials in a fixed order for all subjects. A training trial consisted of briefing by the experimenter, onset of the image, target search and detection by the subject, and indication by the experimenter of the target location. The training trials presented the subject with samples of the type of targets which were used in the experimental set, and backgrounds of comparable complexity and image quality.

The 10 experimental trials were conducted as in the Abstract Scene study, without experimenter feedback on the accuracy of subject responses. When the experimental session was completed, subjects were paid and general information about the experiment was given.

Results and Discussion

The basic data reduction procedure for this experiment was identical to that used in the Abstract Scene experiment. In addition to the sequence of fixation program described previously, an eye fixation scatter plot program was written. This program generated a small symbol at each position fixated during all selected trials. Any set of trials, conditions, and/or subjects could be selected, making it possible to examine the distribution of fixations for an individual subject or image as two examples.

Scatter plots for three images each with two target locations are shown in Figures 16 to 18. In each figure the upper plot shows the distribution of fixations across 10 subjects with the target location the same as used by Scanlan.[54] The lower plot presents the distribution for a second group of 10 subjects with the same background scene but a different target location. Comparison of the distributions between the two target locations reveals that many of the same objects are fixated, although the change in target location causes a change in the variability of the distributions. The data represented by these plots may be used in conjunction with an analysis of the input data in sub-areas of the scene to provide estimates of the perceptual information being extracted.

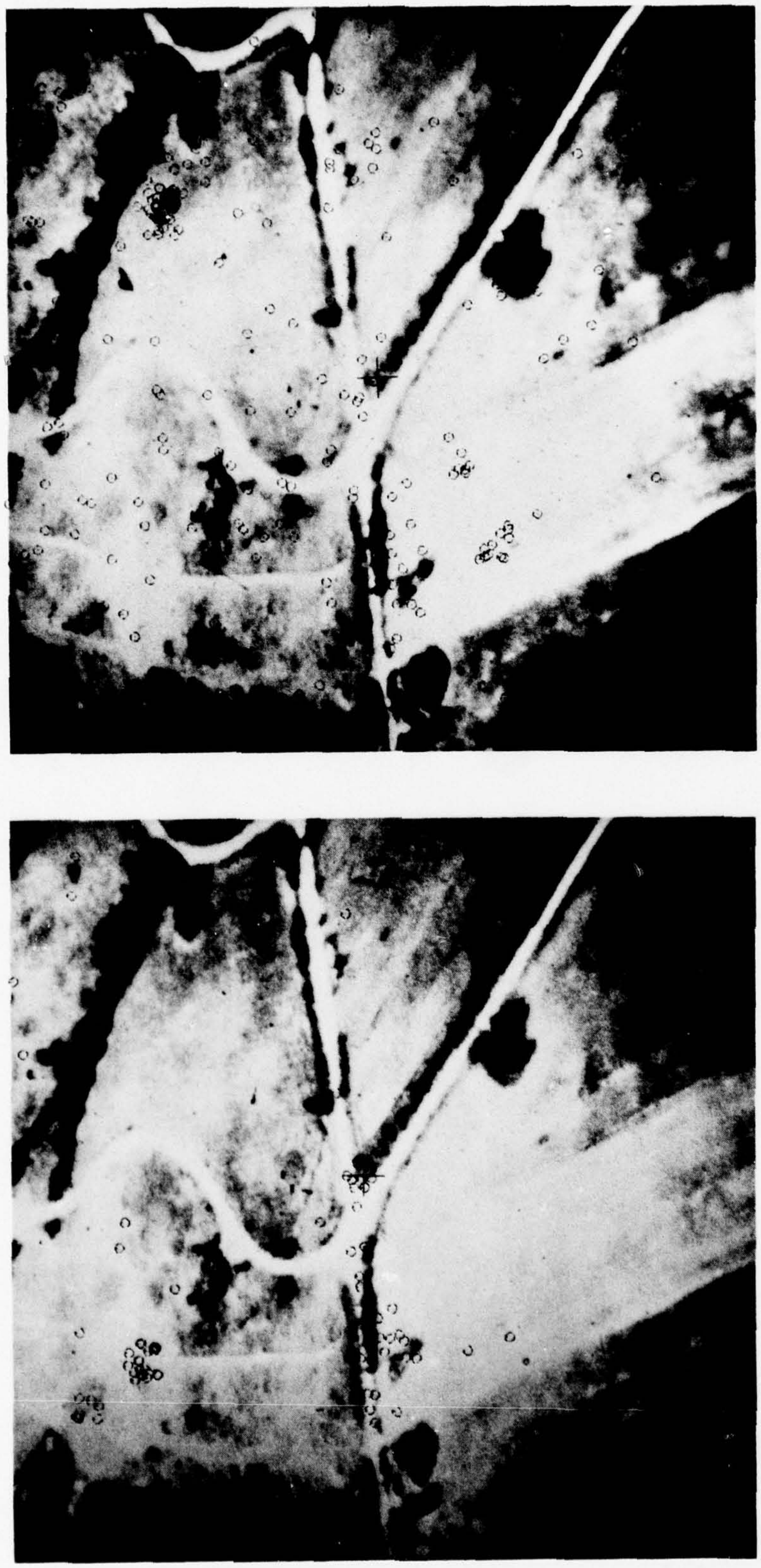


Figure 16. Distribution of fixations on scene Number 0 for 10 subjects.

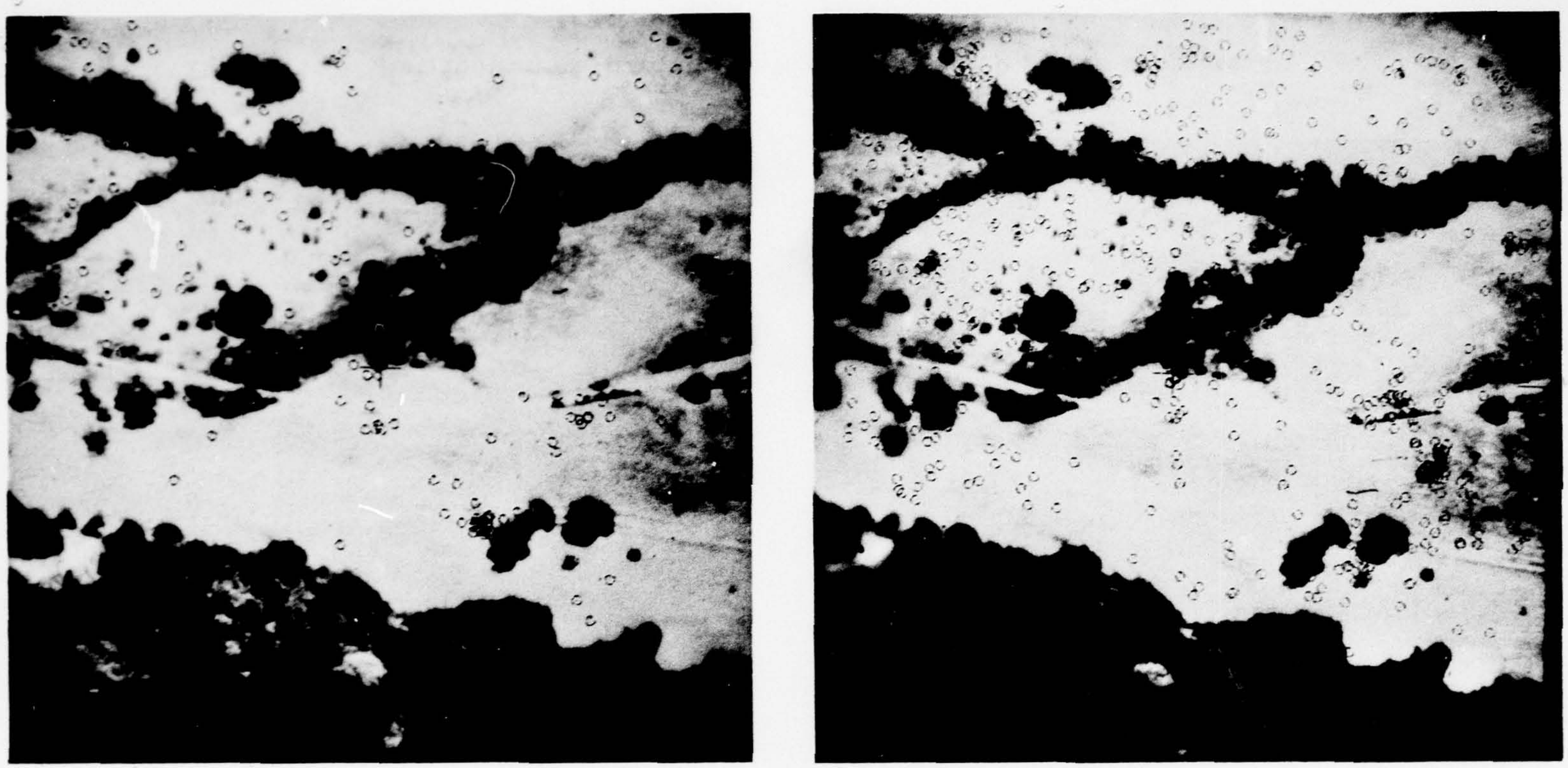


Figure 17. Distribution of fixations on scene Number 3 for 10 subjects.

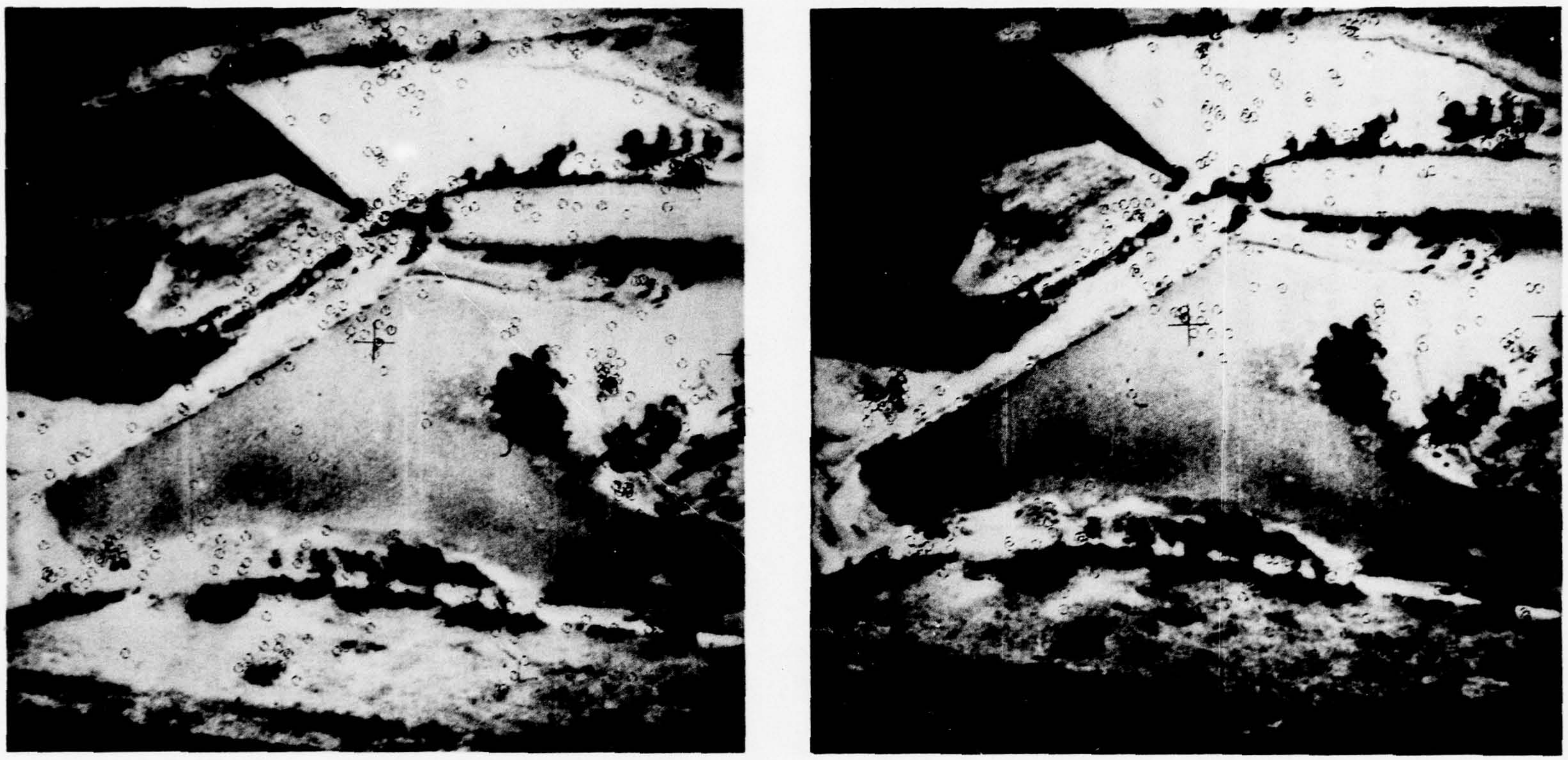


Figure 18. Distribution of fixations on scene Number 7 for 10 subjects.

Input Data. A preliminary data reduction procedure assessed the type of input data available in each image. Two independent raters were presented with the experimental images, over which a seven by seven cell grid was superimposed. The cells covered areas equivalent to approximately 36 mrad on a side, in terms of image subtense when viewed by the subjects. The independent judges used standardized instructions to rate the presence of clutter objects, man-made objects, water, tree masses, roads, and open fields within each sub-area of the image. There was a high degree of agreement between raters and repeated ratings for categories of scene data. Table 1 shows the total number of cells per image within each category.

The target locations within the images were also rated for a context feature called terrain pattern. Terrain pattern categories were based on a literature review of geometric pattern effects on eye fixation behavior. There were six categories in descending order of eye fixation probability. A rating was a function of the highest contrast terrain features within two cells of the target. For example, the target embedded in an image at the point where two roads converged was rated on the terrain pattern made by the road contours. The categories of terrain pattern and their assigned numerical rating in order of decreasing expected probability of fixating were:

1 = target within the vertex of an acute angle.
2 = target within the vertex of angle between 90^{o} and 180^{o}.
3 = target near a straight line.
4 = target in open area, or uniform area.
5 = target near irregular terrain contours (as in tree lines).
6 = target outside the vertex of an angle.

The terrain pattern ratings for the experiment are also given in Table 1.

Fixation Frequency. The proposed Markov process model would estimate the probabilities of transition between states on the basis of the type of information in the scene. Therefore, before the Markov process model can

TABLE 1. NUMBER OF OBJECTS BY CATEGORY OF SCENE DATA, TARGET-TO-BACKGROUND CONTRAST AND TERRAIN PATTERN RATING FOR EACH IMAGE

Image	Target Type	Clutter		Context			Target-to-Background Contrast		Terrain Pattern Rating	
		Number of Objects	Man-Made Objects	Tree Areas	Water Areas	Roads	Target 1	Location 2	Target 1	Location 2
1	APC	20	0	2	0	16	.58	2.06	3	3
2	Truck	28	0	8	0	16	.85	1.90	6	2
3	ABC	30	7	10	0	8	2.39	4.51	4	5
4	Tank	28	0	15	0	12	.63	.02	3	4
5	Tank	27	0	14	1	22	.60	.91	5	1
6	Tank	24	3	18	0	30	1.00	.55	3	4
7	Truck	37	0	10	13	0	.28	.81	4	1
8	Truck	27	0	10	19	23	.73	1.94	6	1
9	Tank	23	6	19	0	23	.17	.42	5	1
10	Truck	27	13	16	5	15	2.72	1.00	3	5

be used, the relationship between eye fixation behavior and input data must be established. One test would be a comparison of frequency of fixation, which may be used to estimate probability of fixation, on specific areas of the images containing objects with features eliciting either search or examination processing.

A first analysis of the relationship between scene input data and perceptual information was accomplished by examining the frequency of fixation on categories of scene data.

One context feature and one clutter feature were selected for preliminary analysis. The presence of any clutter object within a sub-area and the presence of a road or portion of a road in a sub-area were coded for all experimental images by the multiple rater method discussed above. The sub-areas were therefore coded in a 2 by 2 matrix: (00) no clutter objects, no roads (01) no clutter, with roads, (10) clutter objects, no roads, and (11) clutter objects and roads.

Summary scatterplots of all fixations for all subjects viewing each experimental image were generated. The same seven by seven cell rating grid used in the input data assessment procedure was superimposed on the fixation scatterplots, and individual fixations within each of the 49 cells were counted. The cell containing the target was omitted from analysis for all images. The number of fixations for each cell on each image was coded for the clutter and road information conditions, and a total of 480 cells were summarized. The average number of fixations for each clutter-road condition for the 10 experimental images is shown in Table 2.

TABLE 2. AVERAGE FREQUENCY OF FIXATION PER SUB-AREA AS A FUNCTION OF CLUTTER AND ROAD FEATURES

	No Roads	Roads	
No Clutter	2.35	3.30	2.82
Clutter	5.55	8.89	7.22
	3.95	6.09	5.02

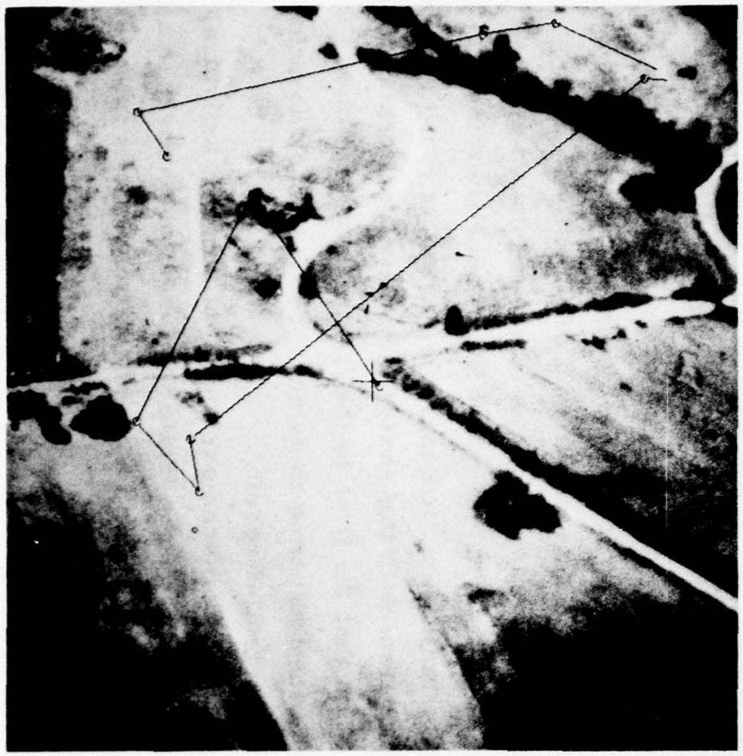
There is a difference in rate of fixation as a function of the presence of clutter and the presence of roads. There is also evidence for an interaction between the two types of input data. A preliminary nonparametric analysis of the data was completed, and an appropriate analysis of variance based on subject-by-subject measures was planned. A Friedman nonparametric analysis[40] showed a reliable effect on clutter-road condition ($p < 0.01$).

Differences in frequency of fixation as a function of information within sub-areas of the scene demonstrates that the operator differentially uses input data during the target search and detection task as hypothesized. Additional analyses examining other types of information and other eye fixation parameters such as duration and sequence will allow a more definitive understanding of the manner in which input data is processed. It is clear, however, that the two processes of search and examination are identifiable and are rotated to the extraction of varying levels of features.

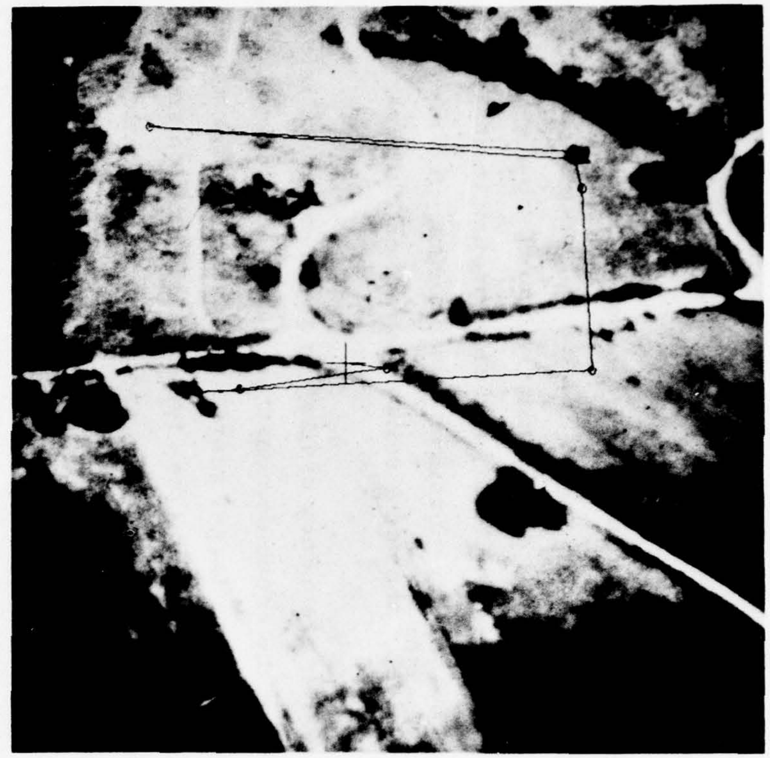
Sequence of Fixations. Plots of the sequence of fixations for trials were also obtained. Figures 19 to 28 present the data for 10 scenes with the two target locations shown in each figure. Figures 29 and 30 present the two images without a target. These data may be used to examine the sequence of fixated objects as well as the probability of fixation. It is anticipated that with suitable measures of scene content and object features, certain commonality in the sequence of fixations will be found.

Informal examination, for example, suggests that the target is often fixated early consistent with the Abstract Scene experimental findings. There is also a suggestion that the time to the first target fixations is much more consistent than is the total time to detect. Further consideration needs to be given to this potential result.

Considerable differences are apparent between individual subjects. Some are unusually fast and can fixate the target very early in a trial. Others require a much larger number of fixations. It is interesting to

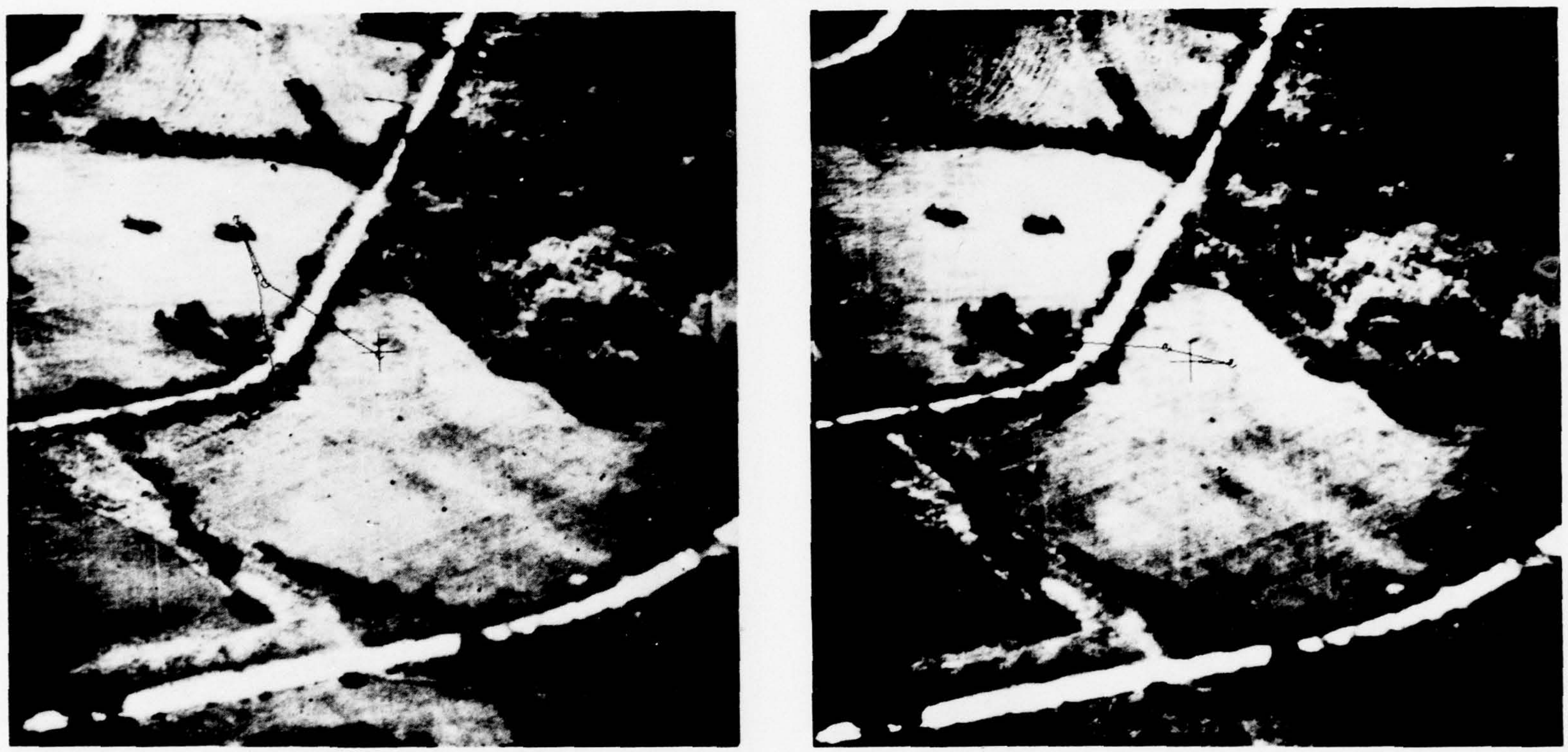


a. Subject A.



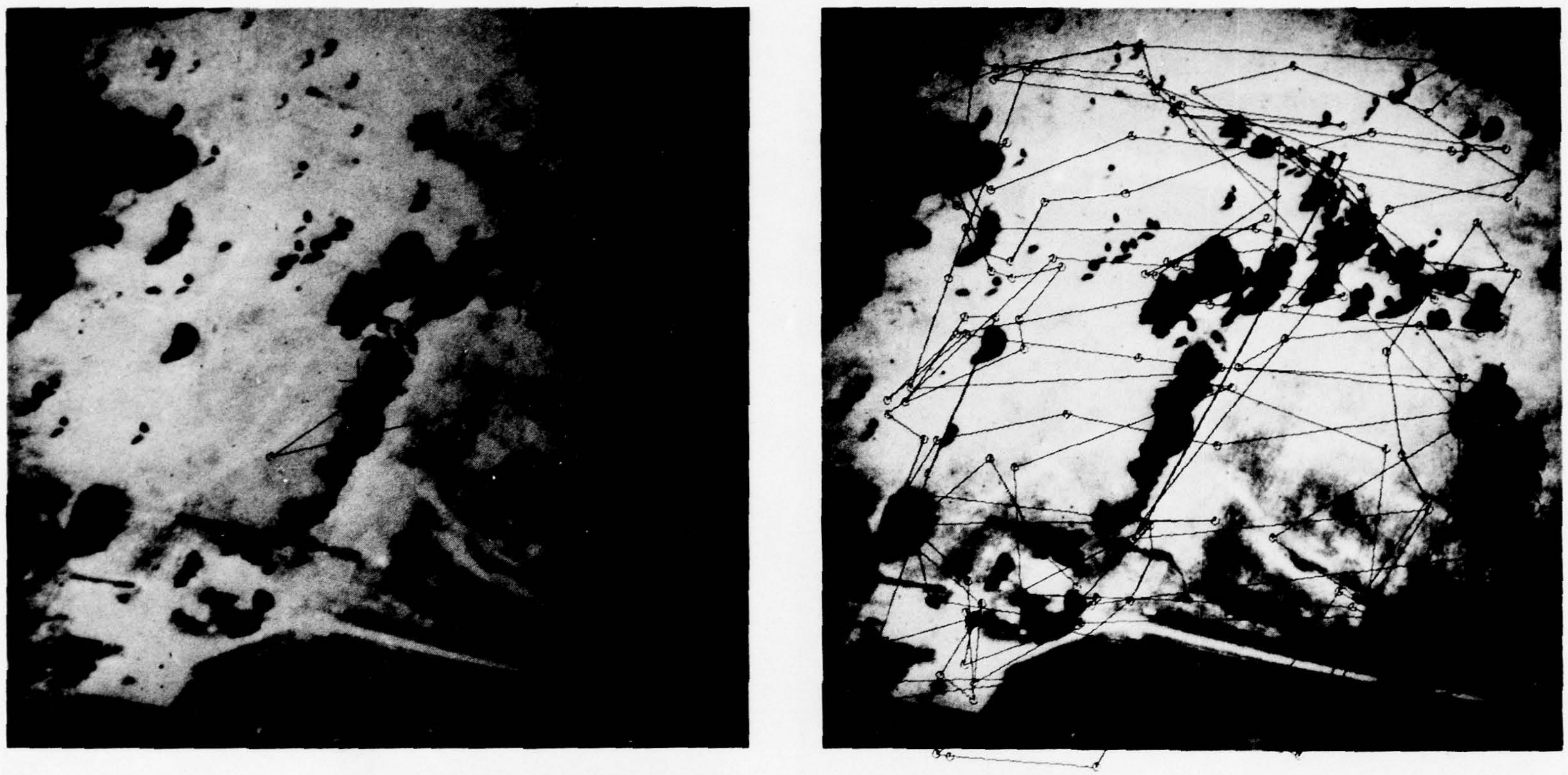
b. Subject B.

Figure 19. Sequence of fixations on scene Number 0 with two different target locations.



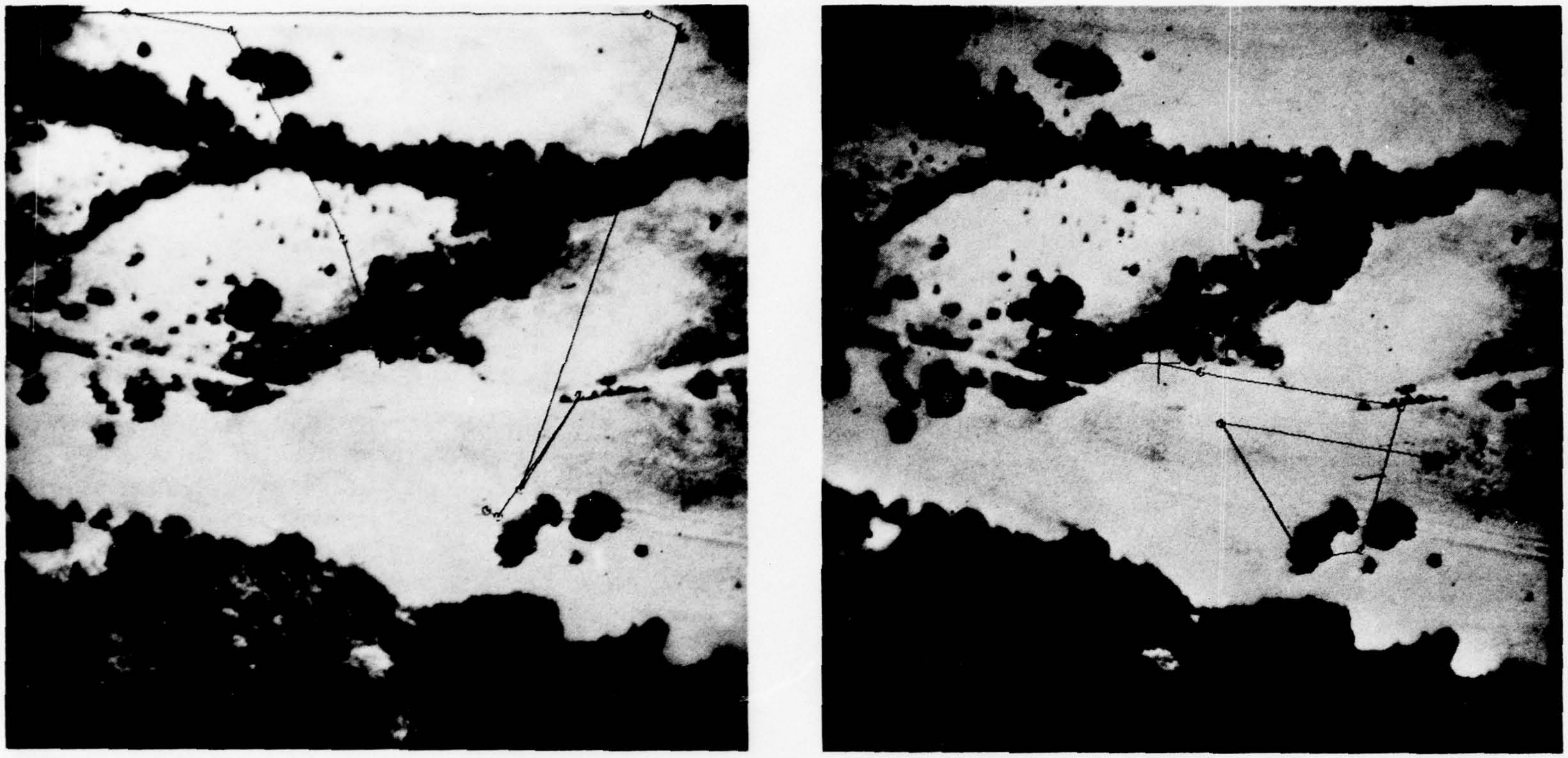
a. Subject A. b. Subject B.

Figure 20. Sequence of fixations on scene Number 1 with two different target locations.



a. Subject A.   b. Subject B.

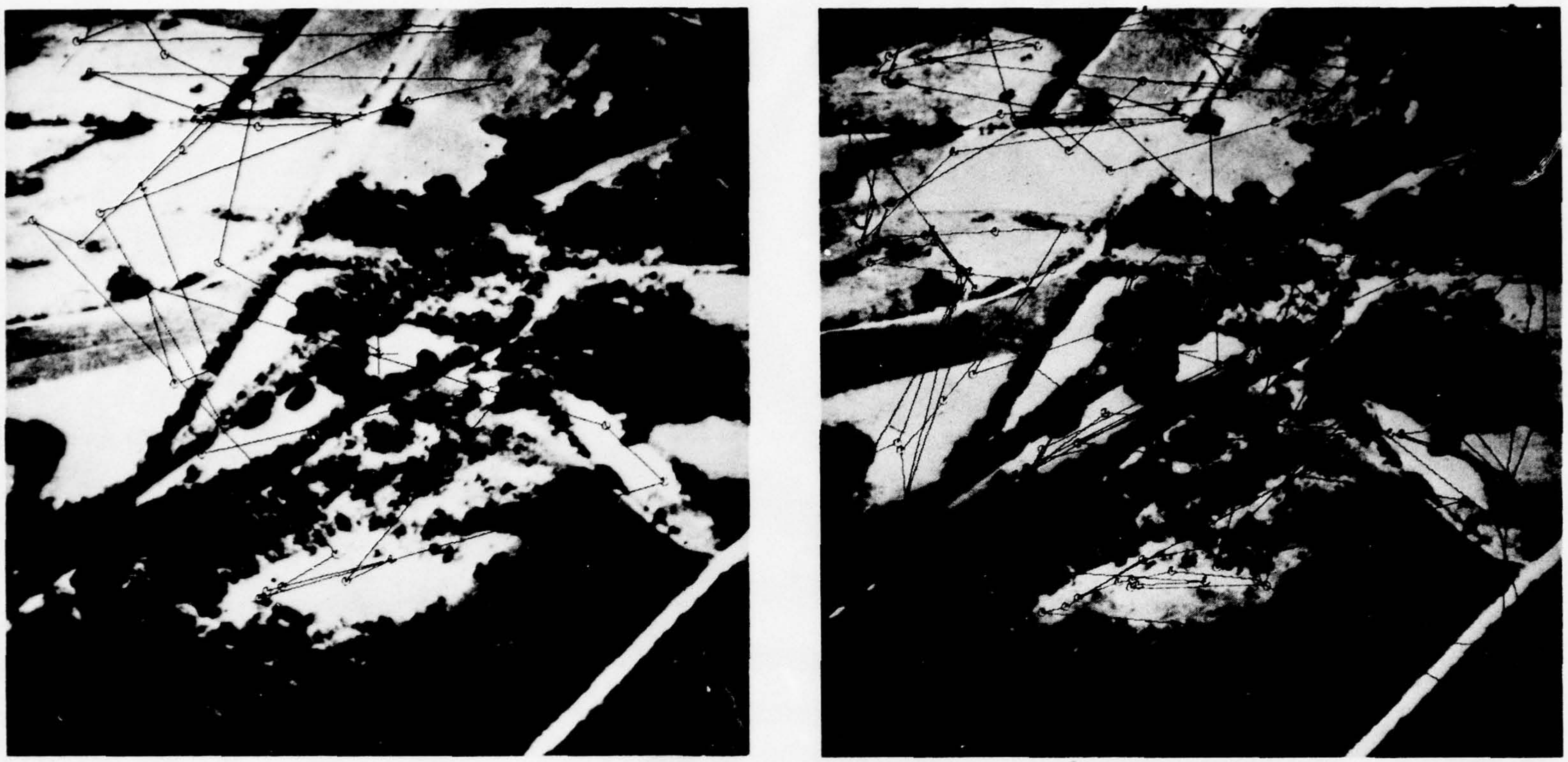
Figure 21. Sequence of fixations on scene Number 2 with two different target locations.



a. Subject A.

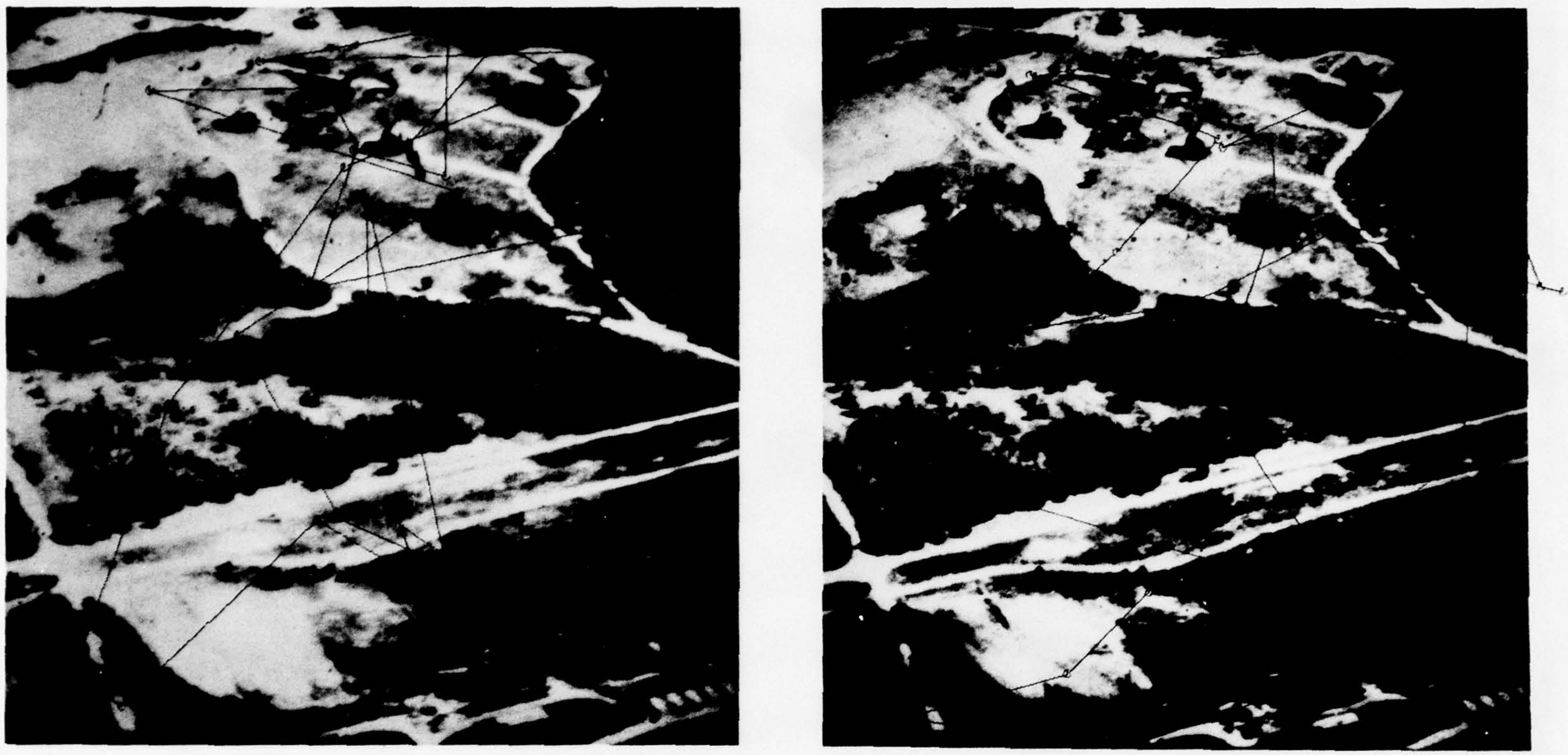
b. Subject B.

Figure 22. Sequence of fixations on scene Number 3 with two different target locations.



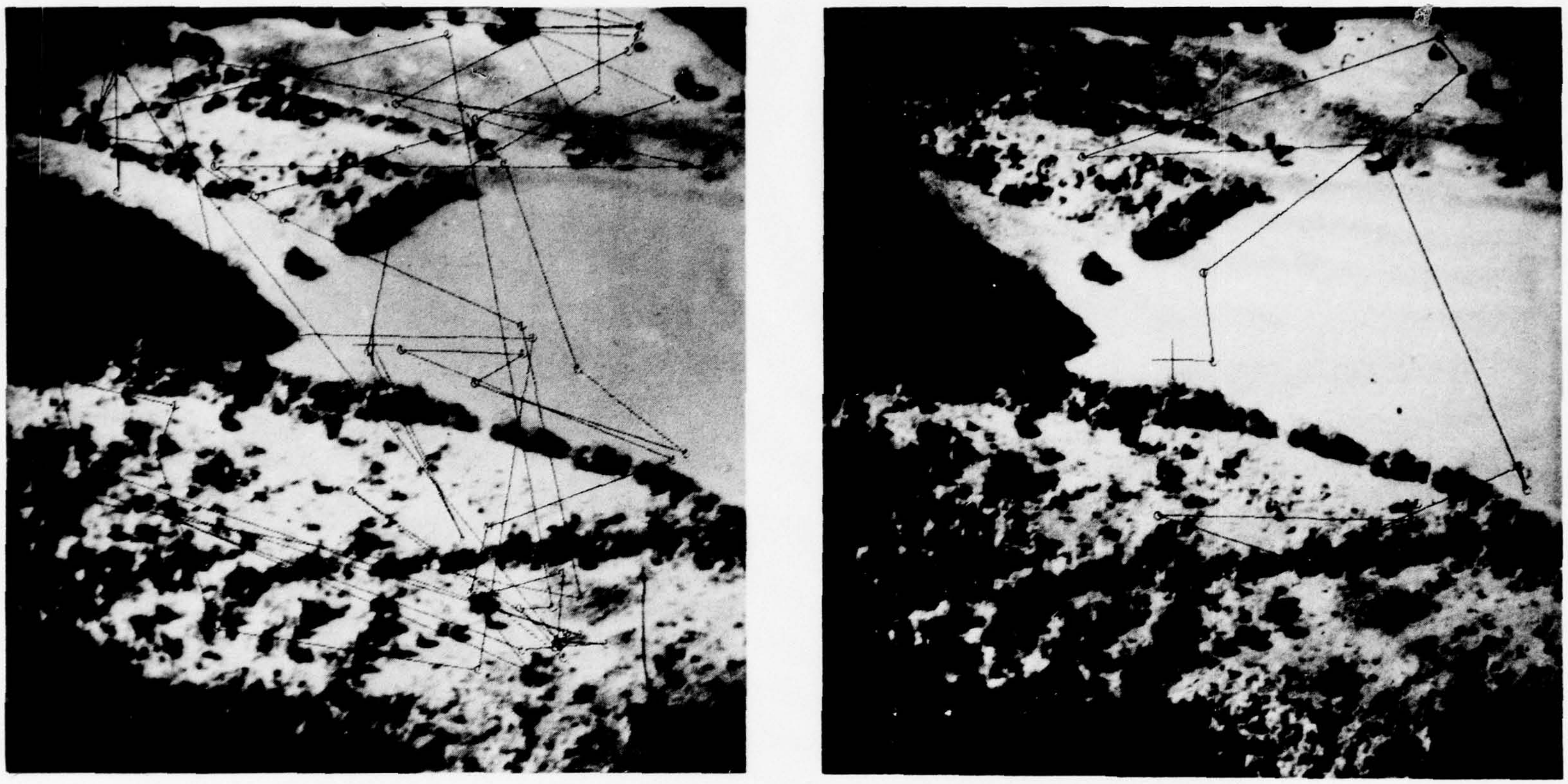
a. Subject A. b. Subject B.

Figure 23. Sequence of fixations on scene Number 4 with two different target locations.



a. Subject A. b. Subject B.

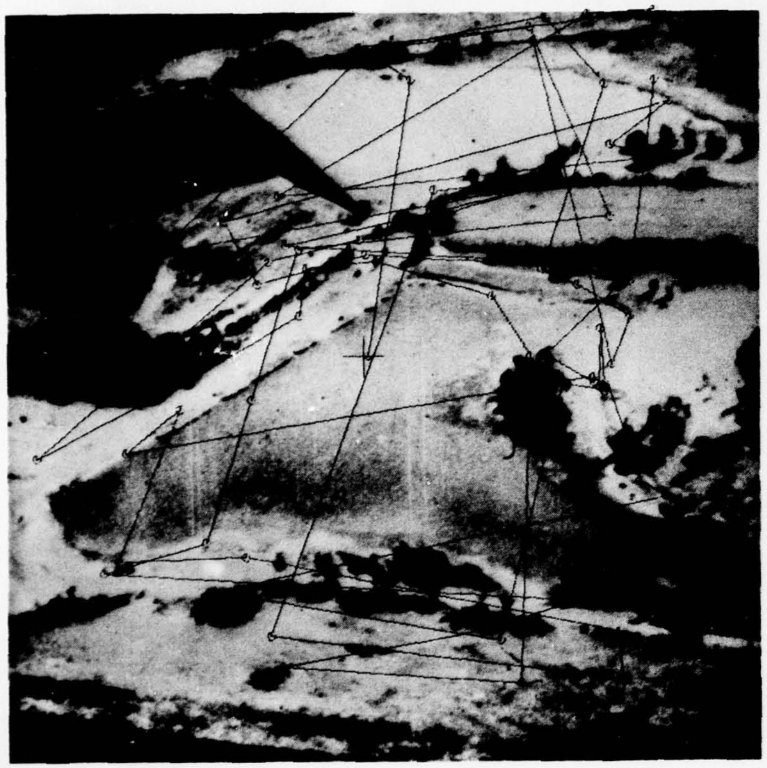
Figure 24. Sequence of fixations on scene Number 5 with two different target locations.



a. Subject A.

b. Subject B.

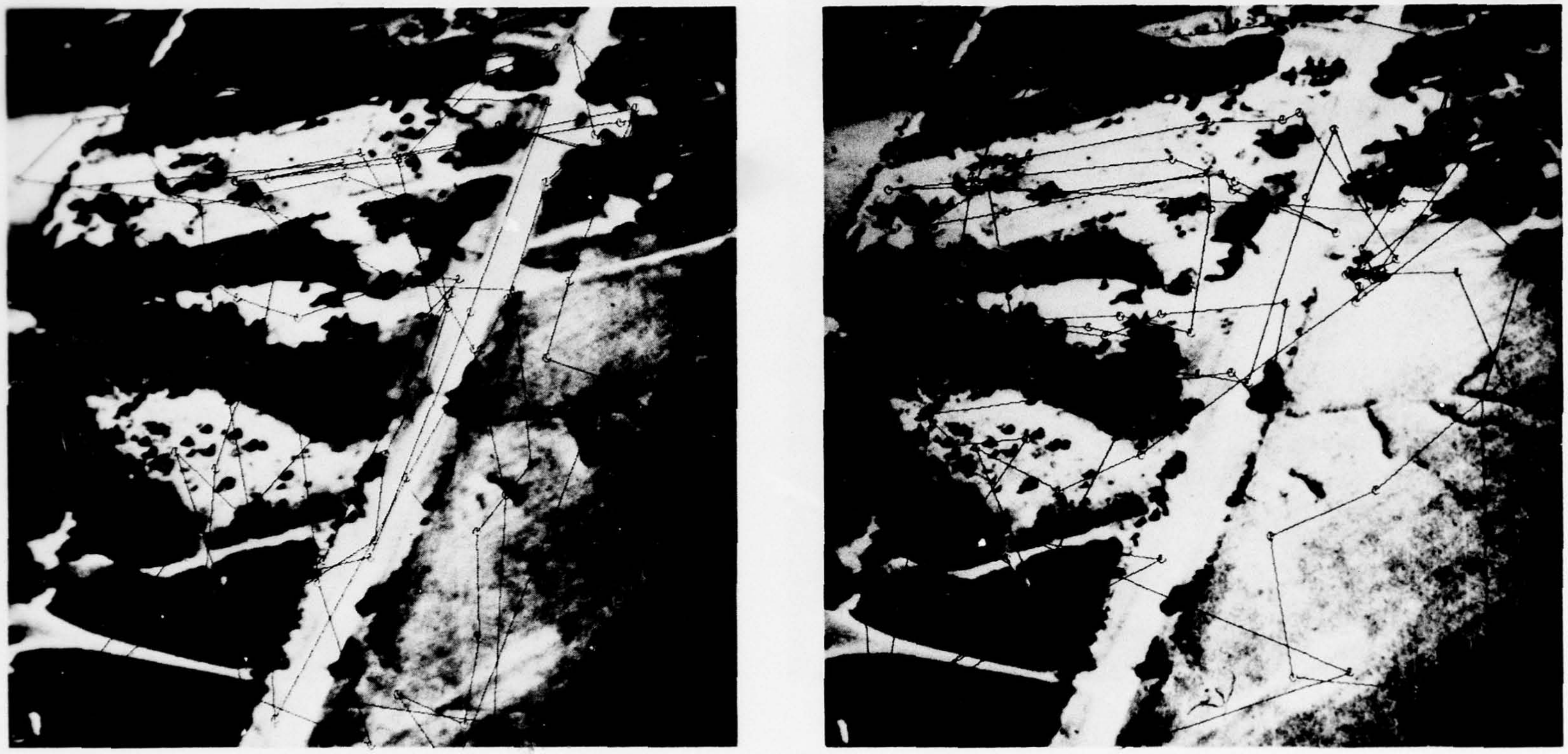
Figure 25. Sequence of fixations on scene Number 6 with two different target locations.



a. Subject A.

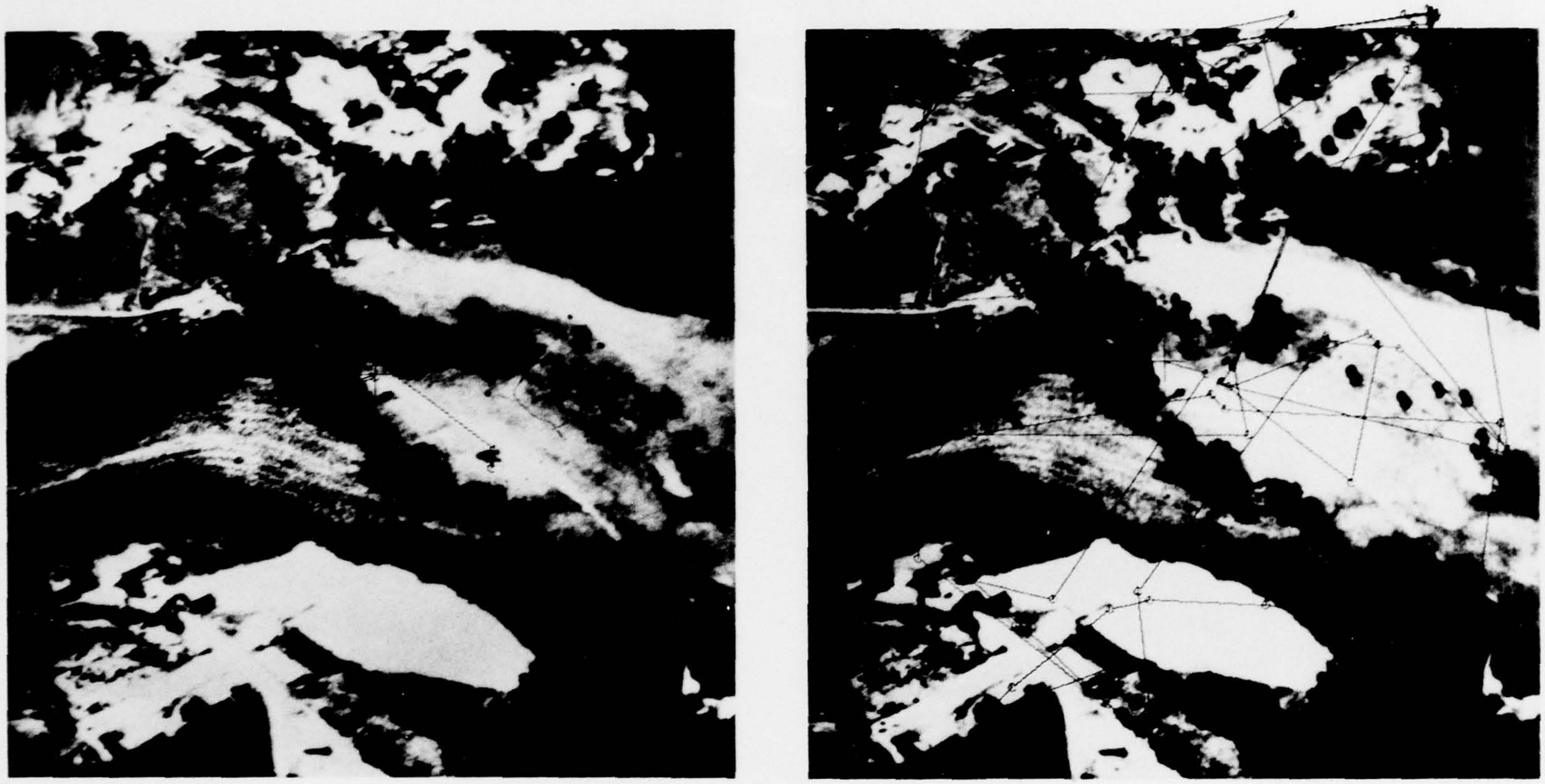
b. Subject B.

Figure 26. Sequence of fixations on scene Number 7 with two different target locations.



a. Subject A. b. Subject B.

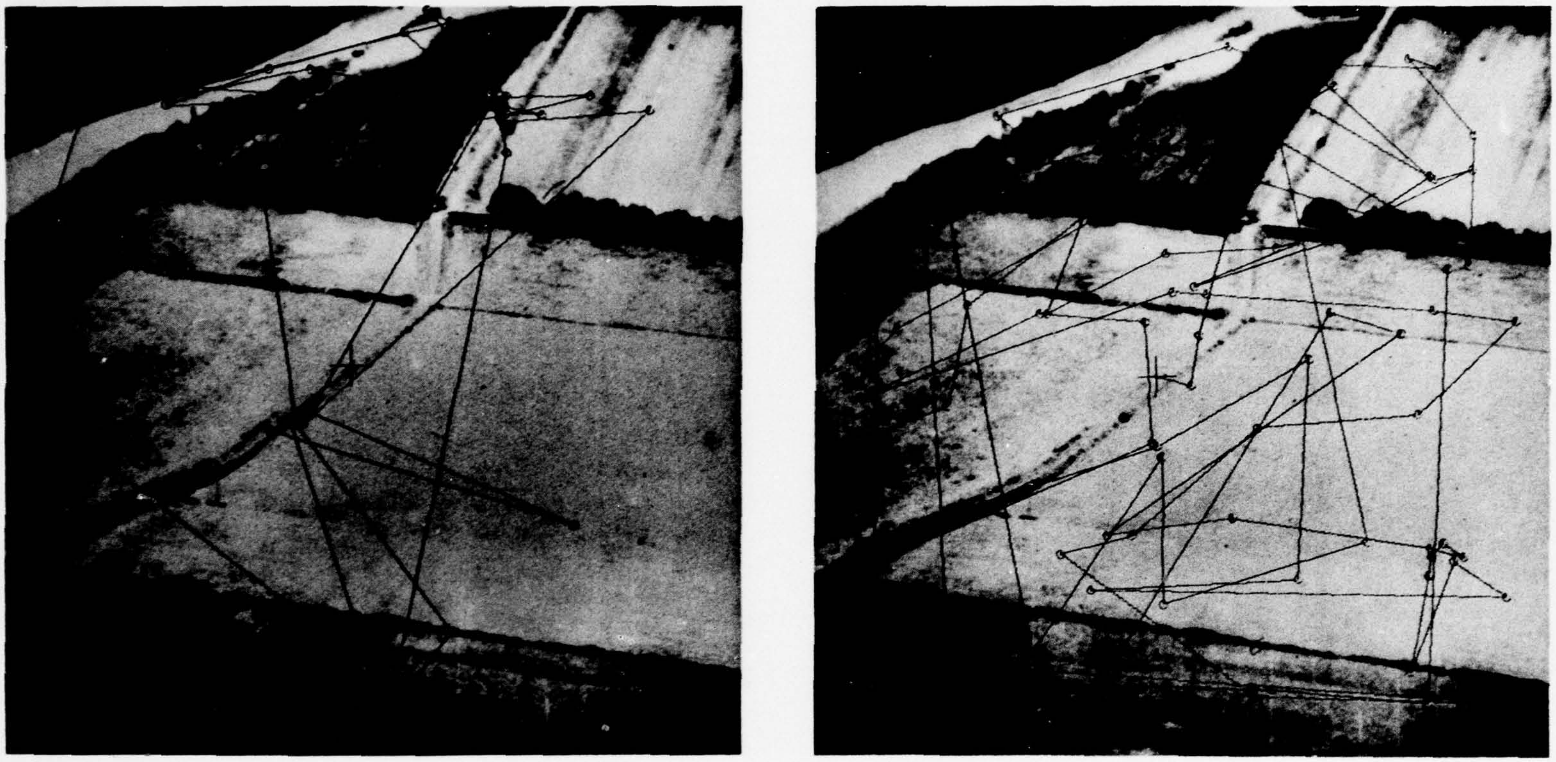
Figure 27. Sequence of fixations on scene Number 8 with two different target locations.



a. Subject A.

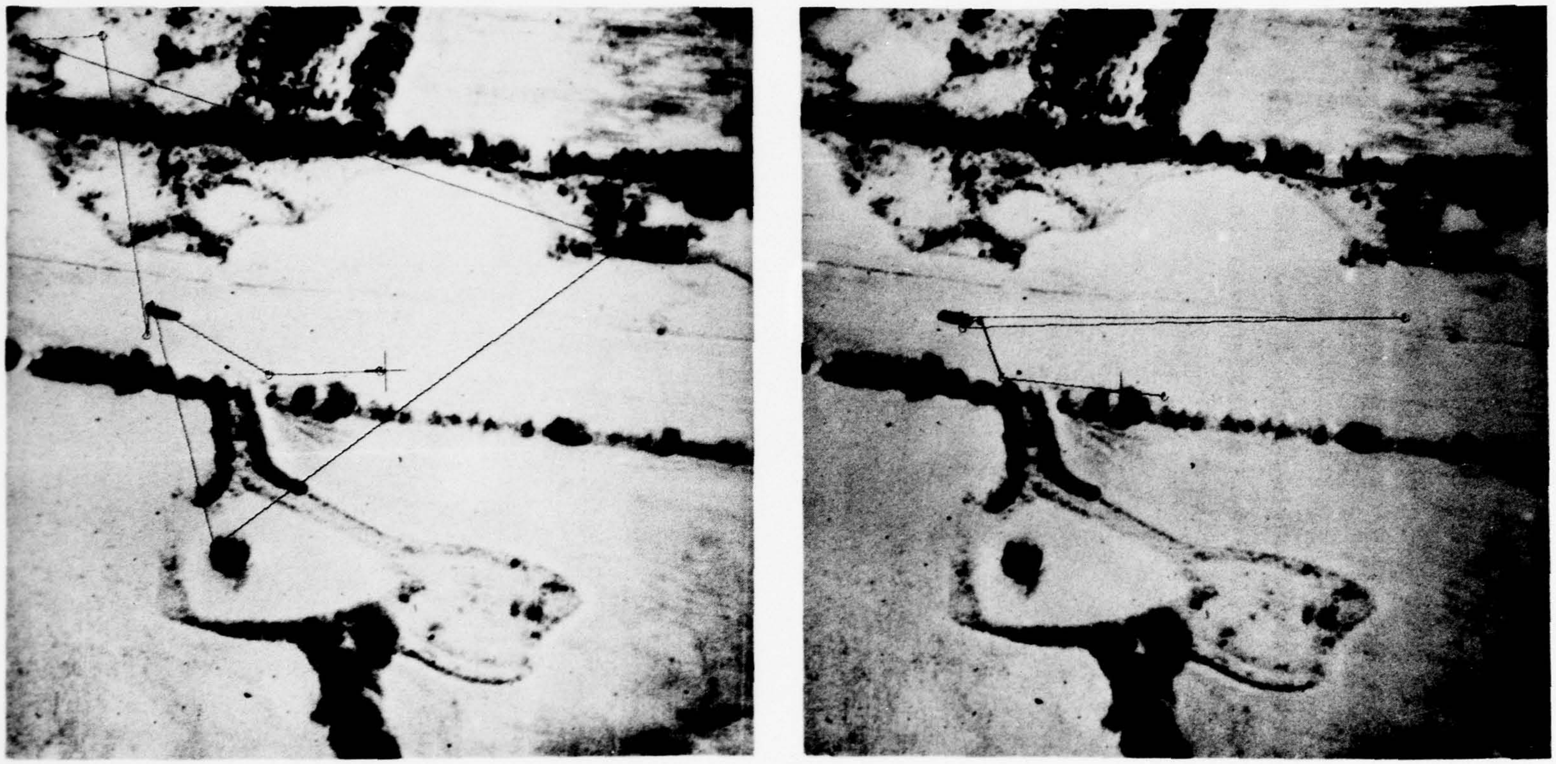
b. Subject B.

Figure 28. Sequence of fixations on scene Number 9 with two different target locations.



a. Subject A. b. Subject B.

Figure 29. Sequence of fixations for two trials without target.



a. Subject A. b. Subject B.

Figure 30. Sequence of fixations for two trials without target.

speculate that such differences may be related to specific prior experiences such as speed reading practice.

Figures 19 to 30 make it abundantly apparent that a scene dependent search strategy is operating. Specific areas of the scene are examined first and some areas are totally ignored. The systematic row by row scan of the scene assumed by many models occurs rarely and then only after failure to detect the target using the scene dependent search.

CONCLUSIONS AND DIRECTIONS

The formulation of target search and detection in terms of the underlying behavioral properties of the observer provides a simple, unified structure on which to build a comprehensive predictive model. This approach creates a framework within which the large body of data accumulated on target and scene characteristics can be organized and incorporated into a mathematical description of search and detection performance over time. In addition, with the use of eye-fixation measures, this formulation allows the generation of specific, testable hypotheses about the information contained in particular characteristics and their function in the search and detection process. The experiments reported were designed to test the validity of these hypotheses in the context of a multi-component information processing model. The preliminary results strongly support the validity of the model through confirmation of key assumptions by means of eye-fixation data, and suggest that a Markov representation will provide a concise, mathematical framework which accurately reflects the underlying behavioral processes.

The Abstract Scene experiment was designed to examine the validity of the assumptions underlying a multi-component process model. In particular, it was hypothesized that the observer selects certain aspects of the input scene for feature analysis, and that the selected aspects will vary according to information need and expectation. It was further proposed that the search and examination states of the model could be characterized by the selected aspects of the input, which were categorized as high and low level features on the basis of spatial frequency components. Eye fixation patterns

were expected to reflect the level of feature processing and thus correspond to the states of the model.

To test the validity of these expectations, high and low level features as well as briefing information were systematically manipulated. The data verified the major assumptions of the model. The influence of a priori expectation was evident both in the patterns of eye fixations and in the probability of target detection over time. In addition, the scan patterns demonstrate that the feature extraction process can be characterized in terms of high and low level analysis, and that the progress of this analysis can be followed over time by examination of the fixation patterns. The multi-component model provides a correspondence between this overt behavior and the underlying information extraction processes. These processes can then be predicted with the Markov representation to ultimately generate a cumulative probability of detection curve.

The results of the Realistic Scene experiment provided additional evidence for the validity of the multi-component model and demonstrated its applicability to realistic scenes. In addition, an abundance of data pertaining to important feature attributes of real scenes and their relation to search strategy was generated. The hypothesized organization of perceptual information into target, clutter, context, and texture results in a useful conceptualization for identifying information requirements for each stage of the target search and detection task. These four categories of scene information also provide a convenient structure for quantifying the content of a scene without resorting to arbitrary metrics. Further, because the influence of each category is different for each stage of the task, the effect of the scene may be more easily incorporated into the model.

The Realistic Scene experiment yielded results which are consistent with those obtained in the abstract scene experiment. The two processes of search and examination were evident from differences in the frequency of fixation as a function of information within sub-areas of the scene. Preliminary analysis of one context feature and one clutter feature, as well as an informal examination of fixation sequences, provided evidence that search strategy is determined by scene content and target context. There was no

evidence that a systematic row by row search strategy was used by any subject, contrary to the assumption used in many search and detection models. However, the presented analysis did not include a consideration of the sequence of fixation in a definitive way. Further examination and consideration of a dominant search pattern, if one exists, will be necessary.

Although a large portion of the data remains to be exploited, the preliminary analyses from both the Abstract and the Realistic Scene experiments indicate that the multi-component, feature extraction model provides a valid and highly useful alternative to the equation-fitting approach. The multi-component behavioral approach integrates and simplifies the large set of potentially relevant scene parameters into generic features. The use of a Markov process offers considerable promise as a model capable of incorporating the indicated processing characteristics in a form that can be expanded as the level of understanding increases. Separation of Search and Examination on the basis of level of feature extraction not only appears warranted but adds significant power to the resulting model. A further expansion of the feature analytic approach to quantifying relevant information has considerable promise as a means of accommodating a large number of system parameters. A change in any system parameter may be described in terms of changes in feature attributes which in turn influence operator performance.

Considerable further analysis must be accomplished to fully exploit the obtained data and to allow elaboration of the reported findings. The limited analysis of variance reported for the Abstract Scene experiment should be expanded to include the effects of individual object features. The dependent measures used in these analyses also need to be expanded to allow an examination of the number of objects fixated as well as the number of fixations. The duration measures used also should be expanded.

The results provide the required data to estimate the transitional probabilities in the Markov model. Accomplishment of this will require a careful definition of the search and examination states as well as an assessment of the extent to which the data meet the assumptions of the model. Through appropriate analysis it will be possible to obtain a set of transition

probabilities for each condition. These can then be used to predict the cumulative probability of detection as a function of time. Discrepancies between actual and predicted results can be used to further refine the model.

The data from the Realistic Scene experiment will provide a wealth of information on the effect of scene content on the information processed by the observer. Both the frequency and sequence of fixations may be used to provide an understanding of scene quantifications. Correlations between scene objects and areas and fixations will indicate the features being habitually used by the observer. Knowledge of the feature attributes of objects can then be combined with knowledge of sensor parameters to relate design parameters to anticipated performance changes.

These data may also be used to test the generalizability of the Markov model derived under the more restricted Abstract Scene conditions. Both sets of data combined and appropriately analyzed will provide a much improved understanding of target search and detection, yield an initial Markov model, and provide the direction necessary for further refinement.

CONCLUDING REMARKS

In summarizing the accomplishments of this phase of the overall Search and Detection modeling effort, one must review them in the context of both previous phase results and the efforts which logically remain to be accomplished. The overall objective of the total program has been to develop an analytical model of the process of search. This model should be capable of predicting the probability of detecting a vehicular type target within a scene which is characteristic of a realistic tactical situation and environment. This probability will be expressed cumulatively as a function of time.

The developmental concept has been to define a simplified model which would have a reasonably high degree of correlation with demonstrable human performance over a wide range of target-background situations. The psycho-physical approach to a performance model appears to offer the highest degree of predictive accuracy among competing approaches and is intuitively satisfying when structured according to Markov modeling theory. It has a high degree of flexibility and, as such, incorporates convenient growth potential.

Because the process by which the human performs an effective search is complex, efforts to understand it and correctly represent the potentially significant factors have led to a relatively complex initial structuring. However we have high confidence that by means of a thorough concluding experimental program, approximations and shortcuts will be defined which will simplify the model without significantly compromising predictive

capability. The overall program has progressed through the following stages thus far:

Phase 1

- Parametric data on target search and detection was provided
- Quantitative scene metries were identified.
- A multi-component process was revealed.

Phase 2

- A multi-component feature analytic model form was developed.
- The Markov mathematical technique was adapted to represent the model.
- Underlying model assumptions were experimentally confirmed.
- Data was provided for parameter estimation.

Future Directions

At this time, the work remaining to be accomplished includes:

Phase 3

- Incorporation of search strategy into the model.
- Determination of model parameter values.
- Validation of model predictions.

Current investigation has shown that the multi-component, feature analytic modeling approach is experimentally valid and most closely represents what is actually happening in the human search process. This understanding and characterization of the fundamental aspects of search represents major progress. Since the extensive preliminary ground work in defining a high accuracy predictive model has now been accomplished, it can be said with confidence that what remains to be done is to flesh out the operating details. Specifically this involves:

- Detailing of the Markov search and examination processes.
- Fitting transition probabilities.
- Incorporating search strategy.

- Incorporating the concept of search area type.
- Providing for Markov expansion and refinement.

The results of the final recommended effort will complete the objectives of the originally proposed three phase program.

REFERENCES

1. Antes, J. R. The time course of picture viewing. Journal of Experimental Psychology, 1974, 103, 67-70.

2. Bailey, H. H. Target acquisition through visual recognition: An early model. Santa Monica, California: Rand Corporation, Technical Report P-4918, September, 1972, (AD-A030 699).

3. Bajcsy, R. and Lieberman, L. L. Computer description of real outdoor scenes. In Proceedings of the Second International Joint Conference on Pattern Recognition. New York: Institute of Electrical and Electronics Engineers, 1974, 174-179.

4. Bajcsy, R. and Rosenthal, D. Visual focusing and defocusing - an essential part of the pattern recognition process. In Proceedings of the Conference on Computer Graphics, Pattern Recognition, and Data Structure of the Institute of Electrical and Electronics Engineers Computer Society. New York: Institute of Electrical and Electronics Engineers, 1975, 130-133.

5. Bajcsy, R. and Tavakoli, M. Computer recognition of roads from satellite pictures. In Proceedings of the Second International Joint Conference on Pattern Recognition. New York: Institute of Electrical and Electronics Engineers, 1974, 190-194.

6. Barlow, H. B., Hill, R. M., and Levick, W. R. Retinal ganglion cells responding selectively to direction and speed of image motion in the rabbit. Journal of Physiology, 1964, 173, 377-407.

7. Von Bekesy, G. Sensory inhibition. Princeton, New Jersey: Princeton University Press, 1967.

8. Bergert, J. W. and Fowler, F. D. Target acquisition studies: Visual angle requirements for TV displayed targets. Orlando, Florida: Martin Marietta Corporation, Technical Report OE10689, May, 1970, (AD-706 369).

9. Biberman, L. M. (Ed.) Perception of displayed information. New York: Plenum Press, 1973.

PRECEDING PAGE BLANK

10. Blesser, B.A., Kuklinski, T.T. and Shillmand, R.J. Empirical tests for features selection based on a psychological theory of character recognition. Pattern Recognition, 1976, 8, 77-85.

11. Box, G.E.P. and Hunter, J.S. Multifactor experimental designs for exploring response surfaces. Annals of Mathematical Statistics, 1957, 28, 195-241.

12. Gox, G.E.P. and Wilson, K.B. On the experimental attainment of optimum conditions. Journal of the Royal Statistical Society, Series B (Methodological), 1950, 13, 1-45.

13. Carterette, E.C. and Friedman, M. (Eds.) Handbook of perception volume V. Seeing. New York: Academic Press, 1975.

14. Cason, G. Image similarity and visual aircraft recognition. Catalogs of Selected Documents in Psychology, 1975, 5, 317.

15. Ciavarelli, A., Walhter, L., and Lee, W. Terrain classification study. China Lake, California: Naval Weapons Center, Technical Report NWC TP-5766, May, 1975, (AD-B005 535L).

16. Clark, C. and Williges, R.C. Response surface methodology central-composite design modifications for human performance research. Human Factors, 1973, 15, 295-310.

17. Cochran, W.G. and Cox, G.M. Experimental designs. New York: Wiley, 1957.

18. Cornsweet, T.N. Visual perception. New York: Academic Press, 1970.

19. Dodwell, P.C. Visual pattern recognition. New York: Holt, Rinehart, and Winston, 1970.

20. Engel, F.V. Visual conspicuity, visual search, and fixation tendencies of the eye. Vision Research, 1977, 17, 95-108.

21. Erickson, R.A. Empirically determined effects of gross terrain features upon ground visibility from low-flying aircraft. China Lake, *California: Naval Ordnance Test Station, Technical Report* NAVWEPS 7779, September, 1961.

22. Fishman, D.H., Hanson, A.R., and Riseman, E.M. Some considerations in a model building system for scene analysis. Amhurst, Massachusetts: University of Massachusetts, Computer and Information Science, Technical Report COINS-TR 75C-2, March, 1975, (AD-A010 149).

AD-A056 760 HUGHES AIRCRAFT CO CULVER CITY CALIF F/G 5/5
A BEHAVIORAL MODEL OF TARGET ACQUISITION IN REALISTIC TERRAIN.(U)
JUN 78 L A SCANLAN, A K AGIN DAAK70-77-C-0013
UNCLASSIFIED HAC-REF-D8983 NL

2 OF 2
AD
A056760

END
DATE
FILMED
9-78
DDC

23. Greening, C. P. Alternative approaches to modeling visual target acquisition. China Lake, California: Naval Weapons Center, Technical Report NWC-TP-5698, September, 1974.

24. Greening, C. P. Mathematical modeling of air-to-ground target acquisition. Human Factors, 1976, 18, 111-148.

25. Haber, R. N. (Ed.) Contemporary theory and research in visual perception. New York: Holt, Rinehart, and Winston, 1968.

26. Huber, R. N. How we remember what we see. Scientific American 1970, 222(5), 104-112.

27. Hall, R. J. and Cusack, B. L. The measurement of eye behavior: Critical and selected reviews of voluntary eye movement and blinking. Aberdeen Proving Ground, Maryland: U. S. Army Human Engineering Laboratory, Technical Memorandum HEL-TM-18-72, July, 1972, (AD-752 904).

28. Hansen, A. R., Riseman, E. M., and Fisher, E. Context in word recognition. Pattern Recognition, 1976, 8, 35-45.

29. Hicks, S. A. and Moler, C. G. A field survey of air-to-ground target detection problems. Aberdeen Proving Ground, Maryland: Human Engineering Laboratory, Technical Memorandum HEL-TM 1-66, January, 1966, (AD 631 361).

30. Hoffman, J. Hierarchical stages in the processing of visual information. Perception and psychophysics, 1975, 18. 348-354.

31. Hubel, D. H., and Wiesel, T. N. Receptive fields; binocular interaction and functional architecture in the cat's visual cortex. Journal of Physiology, 1962, 160, 106-154.

32. Hubel, D. H. and Wiesel, T. N. Receptive fields and functional architecture in two nonstriate visual areas (18 and 19) of the cat. Journal of Neurphysiology, 1965, 28, 228-289.

33. Hubel, D. H. and Wiesel, T. N. Receptive fields and functional architecture of monkey striate cortex. Journal of Physiology (London), 1968, 195, 215-243.

34. Jones, D. B., Freitag, M. and Collyer, S. C. Air-to-ground target acquisition source book: a review of the literature. Orlando, Florida: Martin Marietta Aerospace Company, Technical Report DR 12 470, September, 1974, (AD-A015 079).

35. Julesz, B. Problems of pattern recognition and perceptual psychology. In L. Kanal (Ed.) Proceedings of the 1st IEEE Workshop on Pattern Recognition. Washington, D. C.: Thompson, 1968, 143-154.

36. Kanal, L. (Ed.) Pattern recognition. Washington, D.C.: Thompson, 1968.

37. Kasvand, T. Some observations on linguistics for scene analysis. In Proceedings of the conference on Computer Graphics, Pattern Recognition, and Data Structure of the Institute of Electrical and Electronics Engineers Computer Society. New York: Institute of Electrical and Electronics Engineers, 1975, 118-124.

38. Kirk, R. Experimental design: Procedures for the behavioral sciences. Belmont, California: Brooks-Cole, 1968.

39. Krebs, M. Scanning patterns in real-time FLIR displays. In Hale, R.N., McKnight, R.S., and Moss, J.R. (Eds.) Proceedings of the Nineteenth Annual Meeting of the Human Factors Society. Santa Monica, California: Human Factors Society, October 14-16, 1975, 418-422.

40. Krebs, M. J. and Graf, C. P. Real-time parameters study. Griffiss AFB, New York: Rome Air Development Center, Technical Report RADC TR-73-300, September, 1973, (AD-773 850).

41. Lewis, M.S. Determinants of visual attention in real world scenes. Perceptual and Motor Skills, 1975, 41, 411-416.

42. Lindsay, P.H. and Norman, D.A. Human information processing: An introduction to psychology. New York: Academic Press, 1972.

43. Lipkin, B. and Rosenfeld, A. Picture processing and pschopictorics. New York: Academic Press, 1970.

44. Milner, P.M. A model for visual shape recognition. Psychological Review, 1974, 81, 521-535.

45. Monk, T.H. and Brown, B. The effect of target surround density on visual search performance. Human Factors, 1975, 17, 356-360.

46. Montanari, U. Recent progress in picture analysis. In Proceedings of the Second International Joint Conference on Pattern Recognition. New York: Institute of Electrical and Electronics Engineers, 1974, 513-516.

47. Neisser, U. Cognitive psychology. New York: Appleton-Century-Crofts, 1967.

48. Newman, C.V. Detection of differences between visual textures with varying numbers of dots. Bulletin of the Psychonomic Society, 1974, 4, 201-202.

49. Noton, D. and Stark, L. Eye movements and visual perception. Scientific American, 1971, 224(6), 34-43.

50. Noton, D. and Stark, L. Scanpaths in eye movements during pattern perception. Science, 1971, 171, 308-311.

51. Ratliff, F. Mach bands: Quantitative studies on neural networks in the retina. San Francisco, California: Holden-Day, 1965.

52. Robson, J. G. Receptive fields: Neural representation of the spatial and intensive attributes of the visual images. In E. C. Carterette and M. Friedman (Eds.) Handbook of perception, volume V, Seeing, 1975, 82-116.

53. Rosenfeld, A. Picture processing by computer. New York: Academic Press, 1969.

54. Scanlan, L. A. Target acquisition model development: Effect of realistic terrain. Culver City, California: Hughes Aircraft Company, Displays Systems Department, Man-Machine Systems Section, Technical Report P76-484, December, 1976.

55. Scanlan, L. A. Target acquisition in realistic terrain. In A. S. Neal and R. F. Palasek (Eds.), Proceedings of the Human Factors Society 21st Annual Meeting. Santa Monica, California: Human Factors Society, October, 1977, 249-253.

56. Schnitzler, A. D. Theory of spatial-frequency filtering by the human visual system. I. Performance limited by quantum noise. Journal of the Optical Society of America, 1976, 66, 608-617.

57. Schnitzer, A. D. Theory of spatial-frequency filtering of the human visual system. II. Performance limited by video noise. Journal of the Optical Society of America, 1976, 66, 617-625.

58. Self, H. Image evaluation for the prediction of performance of a human observer. Paper presented at the NATO symposium on Image Evaluation, Kunstlertons, Munich, 1969, 18-22.

59. Selfridge, O. Pandemonium: A paradigm for learning. In Symposium on the mechanization of thought processes. London, England: H. M. Stationery Office, 1959.

60. Simon, C. W. The use of central-composite designs in human factors engineering experiments. Culver City, California: Hughes Aircraft Company, Display Systems and Human Factors Department, Technical Report AFOSR-70-6, December, 1970.

61. Simon, C. W. Economical multifactor designs for human factors engineering experiments. Culver City, California: Hughes Aircraft Company, Display Systems and Human Factors Department, Technical Report P74-541, June, 1973.

62. Smith, S. W. Visual search time and peripheral discriminability. Journal of the Opthamology Society of America, 1961, 51, 14-62.

63. Snyder, H. L. Image Quality and Observer Performance, In L. M. Biberman (Ed.) Perception of displayed information. New York: Plenum Press, 1973.

64. Srinivasan, M. V., Thathachar, M. A., and Deekshatulu, B. L. A probabilistic hypothesis for the prediction of visual fixations. IEEE Transaction of Systems, Man, Cybernetics, 1975, SMC-5, 431-437.

65. Stathacopoulos, A. D., Gilmore, H. F., and Rohringer, G. Review of mathematical models of air-to-ground target acquisition using TV and FLIR sensors. China Lake, California: Naval Weapons Center, Technical Report NWC-TP-5840, January, 1976.

66. Swets, J. A., Tanner, W. P., and Birdsall, T. G. Decision processes in perception. Psychological Review, 1961, 68, 301-340.

67. U.S. Marine Corps tank deployment. Washington, D.C.: U.S. Government Printing Office, Technical Report FMFM9-1, September, 1970.

68. Watanabe, S. Frontiers of pattern recognition. New York: Academic Press, 1972.

69. Weszka, J. and Rosenfeld, A. A comparative study of texture measures for terrain classification. In Proceedings of the conference on Computer Graphics, Pattern Recognition, and Data Structure of the Institute of Electrical and Electronics Engineers Computer Society. New York: Institute of Electrical and Electronics Engineers, 1975, 62-64.

70. White, C. T. and Ford, D. Eye movements during simulated radar search. Journal of the Optical Society of America, 1960, 50, 909-913.

71. Whittenburg, J. A., Schriber, A. L., Robinson, J. D., and Nordlie, P. G. A field study of target detection and identification from the air. Arlington, Virginia: Human Sciences Research, Final Report HSR-TN-59/4 Cue, October, 1959.

72. Williams, L. G. A study of visual search using eye movement recordings. Minneapolis, Minnesota: Honeywell, Inc., Systems and Research Center, Technical Report 12009-IR 2, March, 1967.

73. Williams, L. G., Fairchild, D. D., Graf, C. P., Joula, J. F., and Trumm, G. A. Visual search effectiveness for night vision devices. Minneapolis, Minnesota: Honeywell, Inc. Systems and Research Center, Technical Report 12070-FR, May, 1970.

74. Williges, R. C. and Simon, C. W. Applying response surface methodology to problems of target acquisition. Human Factors, 1971, 13, 511-519.

75. Yarbus, A. L. Eye movements and vision. New York: Plenum Press, 1967.

76. Young, L. R. and Sheena, D. Survey of eye movement recording methods. Behavior Research Methods and Instrumentation, 1975, 7, 397-429.

77. Zavalishin, N. V. Hypothesis concerning the distribution of eye fixation points during the examination of pictures, Automation and Remote Control (USSR), 1968, 12, 1944-1951.

78. Zusne, L. Visual perception of form. New York: Academic Press, 1970.